

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N 15/31, C07K 14/285, 16/12, A61K 31/70, 39/102, 39/40, G01N 33/50		A1	(11) International Publication Number: WO 00/50599
			(43) International Publication Date: 31 August 2000 (31.08.00)
(21) International Application Number: PCT/EP00/01423		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 22 February 2000 (22.02.00)			
(30) Priority Data: 9904183.2 24 February 1999 (24.02.99) GB			
(71) Applicant (for all designated States except US): SMITHKLINE BEECHAM BIOLOGICALS S.A. [BE/BE]; Rue de l'Institut 89, B-1330 Rixensart (BE).			
(72) Inventors; and (75) Inventors/Applicants (for US only): RUELLE, Jean-Louis [BE/BE]; SmithKline Beecham Biologicals s.a., Rue de l'Institut 89, B-1330 Rixensart (BE). THONNARD, Joelle [BE/BE]; SmithKline Beecham Biologicals s.a., Rue de l'Institut 89, B-1330 Rixensart (BE).			
(74) Agent: PRIVETT, Kathryn, Louise; SmithKline Beecham, Two New Horizons Court, Brentford, Middlesex TW8 9EP (GB).		Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(54) Title: HAEMOPHILUS ANTIGEN			
(57) Abstract The invention provides vaccine compositions comprising BASB070 polypeptides and polynucleotides encoding BASB070 polypeptides and methods for producing such polypeptides by recombinant techniques. Also provided are methods for utilizing BASB070 polypeptides to screen for antibacterial compounds.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

HAEMOPHILUS ANTIGEN

FIELD OF THE INVENTION

This invention relates to methods for the production of polynucleotides, (herein referred to as "BASB070" polynucleotide(s)"), polypeptides encoded by them (referred to herein as "BASB070" or "BASB070" polypeptide(s)"), and recombinant materials. In another aspect, the invention relates to methods for using such polypeptides and polynucleotides, including vaccines against bacterial infections. In a further aspect, the invention relates to diagnostic assays for detecting infection of certain pathogens.

BACKGROUND OF THE INVENTION

Haemophilus influenza is a non-motile Gram negative bacterium. Man is its only natural host. *H. influenzae* isolates are usually classified according to their polysaccharide capsule. Six different capsular types designated a through f have been identified. Isolates that fail to agglutinate with antisera raised against one of these six serotypes are classified as nontypeable, and do not express a capsule.

The *H. influenzae* type b is clearly different from the other types in that it is a major cause of bacterial meningitis and systemic diseases. Nontypeable *H. influenzae* (NTHi) are only occasionally isolated from the blood of patients with systemic disease.

NTHi is a common cause of pneumonia, exacerbation of chronic bronchitis, sinusitis and otitis media.

Otitis media is an important childhood disease both by the number of cases and its potential sequelae. More than 3.5 millions cases are recorded every year in the United

States, and it is estimated that 80 % of children have experienced at least one episode of otitis before reaching the age of 3 (1). Left untreated, or becoming chronic, this disease may lead to hearing loss that can be temporary (in the case of fluid accumulation in the middle ear) or permanent (if the auditive nerve is damaged). In infants, such hearing
5 losses may be responsible for delayed speech learning.

Three bacterial species are primarily isolated from the middle ear of children with otitis media: *Streptococcus pneumoniae*, NTHi and *M. catarrhalis*. These are present in 60 to 90 % of cases. A review of recent studies shows that *S. pneumoniae* and NTHi together
10 represent about 30 %, and *M. catarrhalis* about 15 % of otitis media cases (2). Other bacteria can be isolated from the middle ear (*H. influenza* type B, *S. pyogenes*, ...) but at a much lower frequency (2 % of the cases or less).

Epidemiological data indicate that, for the pathogens found in the middle ear, the
15 colonization of the upper respiratory tract is an absolute prerequisite for the development of an otitis; other factors are however also required to lead to the disease (3-9). These are important to trigger the migration of the bacteria into the middle ear via the Eustachian tubes, followed by the initiation of an inflammatory process. These other factors are unknown todate. It has been postulated that a transient anomaly of the immune system
20 following a viral infection, for example, could cause an inability to control the colonization of the respiratory tract (5). An alternative explanation is that the exposure to environmental factors allows a more important colonization of some children, who subsequently become susceptible to the development of otitis media because of the sustained presence of middle ear pathogens (2).

25

Various proteins of *H. influenzae* have been shown to be involved in pathogenesis or have been shown to confer protection upon vaccination in animal models.

Adherence of NTHi to human nasopharygeal epithelial cells has been reported (10).

Apart from fimbriae and pili (11-15), many adhesins have been identified in NTHi.

Among them, two surface exposed high-molecular-weight proteins designated HMW1 and HMW2 have been shown to mediate adhesion of NTHi to epithelial cells (16).

- 5 Another family of high molecular weight proteins has been identified in NTHi strains that lack proteins belonging to HMW1/HMW2 family. The NTHi 115 kDa Hia protein (17) is highly similar to the Hsf adhesin expressed by *H. influenzae* type b strains (18). Another protein, the Hap protein shows similarity to IgA1 serine proteases and has been shown to be involved in both adhesion and cell entry (19).

10

Five major outer membrane proteins (OMP) have been identified and numerically numbered.

- 15 Original studies using *H. influenzae* type b strains showed that antibodies specific for P1 and P2 protected infant rats from subsequent challenge (20-21). P2 was found to be able to induce bactericidal and opsonic antibodies, which are directed against the variable regions present within surface exposed loop structures of this integral OMP (22-23). The lipoprotein P4 also could induce bactericidal antibodies (24).

- 20 P6 is a conserved peptidoglycan-associated lipoprotein making up 1-5 % of the outer membrane (25). Later a lipoprotein of about the same mol. wt. was recognized, called PCP (P6 crossreactive protein) (26). A mixture of the conserved lipoproteins P4, P6 and PCP did not reveal protection as measured in a chinchilla otitis-media model (27). P6 alone appears to induce protection in the chinchilla model (28).

25

Another fimbrin is described with homology to P5, which in itself has sequence homology to the integral *Escherichia coli* OmpA (29-30). This paradox needs further investigation to clarify the nature and role of pilin, pilin-associated proteins, pilin-

excreting proteins and P5. It is however shown that NTHi adhere to mucus by way of fimbriae. (29). P5 appears to undergo antigenic drift during persistent infections with NTHi (31).

- 5 In line with the observations made with gonococci and meningococci, NTHi expresses a dual human transferrin receptor composed of TbpA and TbpB when grown under iron limitation. Anti-TbpB protected infant rats. (32). Hemoglobin / haptoglobin receptors have also been described for NTHi (33). A receptor for Haem: Hemopexin has also been identified (34). A lactoferrin receptor is also present in NTHi, but is not yet characterized
- 10 (35). A protein resembling neisserial FrpB-protein has not been described in NTHi.

- A 80kDa OMP, the D15 surface antigen, provides protection against NTHi in a mouse challenge model. (36). A 42kDa outer membrane lipoprotein, LPD is conserved amongst *Haemophilus influenzae* and induces bactericidal antibodies (37). A minor 98kDa OMP
- 15 (38), was found to be a protective antigen, this OMP may very well be one of the Fe-limitation inducible OMPs or high molecular weight adhesins that have been characterized thereafter. *H. Influenzae* produces IgA1-protease activity (39). IgA1-proteases of NTHi reveals a high degree of antigenic variability (40).
- Another OMP of NTHi, OMP26, a 26-kDa protein has been shown to enhance
- 20 pulmonary clearance in a rat model (41). The NTHi HtrA protein has also been shown to be a protective antigen. Indeed, this protein protected Chinchilla against otitis media and protected infant rats against *H. influenzae* type b bacteremia (42)

Background References

- 25 1. Klein, JO (1994) Clin.Inf.Dis 19:823
2. Murphy, TF (1996) Microbiol.Rev. 60:267

3. Dickinson, DP et al. (1988) *J. Infect.Dis.* 158:205
4. Faden, HL et al. (1991) *Ann.Otorhinol.Laryngol.* 100:612
5. Faden, HL et al (1994) *J. Infect.Dis.* 169:1312
6. Leach, AJ et al. (1994) *Pediatr.Infect.Dis.J.* 13:983
- 5 7. Prellner, KP et al. (1984) *Acta Otolaryngol.* 98:343
8. Stenfors, L-E and Raisanen, S. (1992) *J.Infect.Dis.* 165:1148
9. Stenfors, L-E and Raisanen, S. (1994) *Acta Otolaryngol.* 113:191
10. Read, RC. et al. (1991) *J. Infect. Dis.* 163:549
11. Brinton, CC. et al. (1989) *Pediatr. Infect. Dis. J.* 8:S54
- 10 12. Kar, S. et al. (1990) *Infect. Immun.* 58:903
13. Gildorf, JR. et al. (1992) *Infect. Immun.* 60:374
14. St. Geme, JW et al. (1991) *Infect. Immun.* 59:3366
15. St. Geme, JW et al. (1993) *Infect. Immun.* 61: 2233
16. St. Geme, JW. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:2875
- 15 17. Barenkamp, SJ. et JW St Geme (1996) *Mol. Microbiol.* (In press)
18. St. Geme, JW. et al. (1996) *J. Bact.* 178:6281
19. St. Geme, JW. et al. (1994) *Mol. Microbiol.* 14:217
20. Loeb, MR. et al. (1987) *Infect. Immun.* 55:2612
21. Musson, RS. Jr. et al. (1983) *J. Clin. Invest.* 72:677
- 20 22. Haase, EM. et al. (1994) *Infect. Immun.* 62:3712
23. Troelstra, A. et al. (1994) *Infect. Immun.* 62:779
24. Green, BA. et al. (1991) *Infect.Immun.*59:3191
25. Nelson, MB. et al. (1991) *Infect. Immun.* 59:2658
26. Deich, RM. et al. (1990) *Infect. Immun.* 58:3388
- 25 27. Green, BA. et al. (1993) *Infect.immun.* 61:1950
28. Demaria, TF. et al. (1996) *Infect. Immun.* 64:5187
29. Miyamoto, N., Bakaletz, LO (1996) *Microb. Pathog.* 21:343
30. Munson, RS.j.r. et al. (1993) *Infect. Immun.* 61:1017

31. Duim, B. et al. (1997) *Infect. Immun.* 65:1351
32. Loosmore, SM. et al (1996) *Mol. Microbiol.* 19:575
33. Maciver, I. et al. (1996) *Infect. Immun.* 64:3703
34. Cope, LD. et al. (1994) *Mol. Microbiol.* 13:868
- 5 35. Schryvers, AB. et al. (1989) *J. Med. Microbiol.* 29:121
36. Flack, FS. et al. (1995) *Gene* 156:97
37. Akkoyunlu, M. et al. (1996) *Infect. Immun.* 64:4586
38. Kimura, A. et al. (1985) *Infect. Immun.* 47:253
39. Mulks, MH. et Shoberg, RJ (1994) *Meth. Enzymol.* 235:543
- 10 40. Lomholt, H. Alphen, Lv, Kilian, M. (1993) *Infect. Immun.* 61:4575
41. Kyd, J.M. and Cripps, A.W. (1998) *Infect. Immun.* 66:2272
42. Loosmore, S.M. et al. (1998) *Infect. Immun.* 66:899

The frequency of NTHi infections has risen dramatically in the past few decades. This
15 phenomenon has created an unmet medical need for new anti-microbial agents, vaccines,
drug screening methods and diagnostic tests for this organism. The present invention
aims to meet that need. In particular the present invention aims to meet the need for a
vaccine effective against NTHi.

20

SUMMARY OF THE INVENTION

The present invention relates to recombinant materials and methods for the production of
BASB070, in particular BASB070 polypeptides and BASB070 polynucleotides, for use
25 especially in therapeutic or prophylactic vaccines. In another aspect, the invention relates to
methods for using such polypeptides and polynucleotides, including prevention and
treatment of microbial diseases, amongst others. In a further aspect, the invention relates to
diagnostic assays for detecting diseases associated with microbial infections and conditions

associated with such infections, such as assays for detecting expression or activity of BASB070 polynucleotides or polypeptides.

5 It has been discovered that BASB070 encodes a polypeptide that has the features of a surface-exposed molecule recognisable by the immune system. For example, the polypeptide encoded by BASB070 contains a signal peptide, indicating that it is exported at least to the periplasm between the inner and outer membranes of the bacterium. Furthermore the polypeptide has similarities to other known surface-exposed proteins and potential similarity to other known immunogenic and immunoprotective peptides.

10 BASB070 is 26% identical to the HasR protein of *Serratia marcescens* in an 817 amino acid overlap. *S. marcescens* HasR is a receptor for the HasA hemophore protein. It is a TonB dependent protein. It has the characteristics of an integral outer membrane protein with a β -barrel 3D structure. The β -barrels formed by the integral outer membrane
15 proteins are composed of anti-parallel, amphipathic β -strands. Their external loops contain frequently immunodominant B-cell epitopes. BASB070 is sufficiently closely related to the HasR protein of *Serratia marcescens* to say that BASB070 is also an integral outer membrane protein with a β -barrel conformation. BASB070 or fragments of it therefore provide potential vaccine antigens.

20 Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following descriptions and from reading the other parts of the present disclosure.

25

DESCRIPTION OF THE INVENTION

The invention relates to the use of BASB070 polypeptides and polynucleotides as described in greater detail below. In particular, the invention relates to the use of polypeptides and polynucleotides of a BASB070 of *Haemophilus influenzae*, which is related by amino acid sequence homology to *Serratia marcescens* HasR hemophore receptor polypeptide. The invention relates especially to the use of BASB070 having the nucleotide and amino acid sequences set out in SEQ ID NO:1 or 3 and SEQ ID NO:2 or 4 respectively.

The invention further relates to uses of polynucleotides and polypeptides which have at least 85% identity, preferably at least 90% identity, more preferably at least 95% identity, most preferably at least 97-99% or exact identity to the sequences identified in SEQ ID NO:1 or 3 and SEQ ID NO:2 or 4.

The invention also relates to novel NTHi polynucleotide and polypeptide sequences disclosed herein.

Polypeptides

In one aspect of the invention there are provided uses for polypeptides of *Haemophilus influenzae* referred to herein as "BASB070" and "BASB070 polypeptides" as well as biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

The present invention further provides uses for:

- (a) an isolated polypeptide which comprises an amino acid sequence which has at least 85% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% or exact identity, to that of SEQ ID NO:2 or 4;
- (b) a polypeptide encoded by an isolated polynucleotide comprising a polynucleotide sequence which has at least 85% identity, more preferably at least 90% identity, yet more

preferably at least 95% identity, even more preferably at least 97-99% or exact identity to SEQ ID NO:1 or 3 over the entire length of SEQ ID NO:1 or 3 respectively; or

- (c) a polypeptide encoded by an isolated polynucleotide comprising a polynucleotide sequence encoding a polypeptide which has at least 85% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% or exact identity, to the amino acid sequence of SEQ ID NO:2 or 4.

The BASB070 polypeptides provided in SEQ ID NO:2 or 4 are the BASB070 polypeptides from *Haemophilus influenzae* strains Rd KW20 and ntHi 3224.

10

- The invention also provides uses for immunogenic fragments of a BASB070 polypeptide, that is, a contiguous portion of the BASB070 polypeptide which has the same or substantially the same immunogenic activity as the polypeptide comprising the amino acid sequence of SEQ ID NO:2 or 4. That is to say, the fragment (if necessary when coupled to a carrier) is capable of raising an immune response which recognises the BASB070 polypeptide. Such an immunogenic fragment may include, for example, the BASB070 polypeptide lacking an N-terminal leader sequence, and/or a transmembrane domain and/or a C-terminal anchor domain. In a preferred aspect the immunogenic fragment of BASB070 according to the invention comprises substantially all of the extracellular domain of a polypeptide which has at least 85% identity, preferably at least 90% identity, more preferably at least 95% identity, most preferably at least 97-99% identity, more preferably at least more preferably at least 95% identity, most preferably at least 97-99% identity, to that of SEQ ID NO:2 or 4 over the entire length of SEQ ID NO:2 or 4.

- 25 A fragment is a polypeptide having an amino acid sequence that is entirely the same as part but not all of any amino acid sequence of any polypeptide of the invention. As with BASB070 polypeptides, fragments may be "free-standing", or comprised within a larger

polypeptide of which they form a part or region, most preferably as a single continuous region in a single larger polypeptide.

5 Preferred fragments include, for example, truncation polypeptides having a portion of an amino acid sequence of SEQ ID NO:2 or 4 or of a variant thereof, such as a continuous series of residues that includes an amino- and/or carboxyl-terminal amino acid sequence. Degradation forms of the polypeptides of the invention produced by or in a host cell, are also preferred. Further preferred are fragments characterized by structural or functional attributes such as fragments that comprise beta-barrels, alpha-helix and alpha-helix forming
10 regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions.

15 Further preferred fragments include an isolated polypeptide comprising an amino acid sequence having at least 15, 20, 30, 40, 50 or 100 contiguous amino acids from the amino acid sequence of SEQ ID NO:2 or 4, or an isolated polypeptide comprising an amino acid sequence having at least 15, 20, 30, 40, 50 or 100 contiguous amino acids truncated or deleted from the amino acid sequence of SEQ ID NO:2 or 4.

20

Particularly preferred are variants in which several, 5-10, 1-5, 1-3, 1-2 or 1 amino acids are substituted, deleted, or added in any combination.

25 The polypeptides, or immunogenic fragments, for use in the invention may be in the form of the "mature" protein or may be a part of a larger protein such as a precursor or a fusion protein. It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or an additional sequence for stability

during recombinant production. Furthermore, addition of exogenous polypeptide or lipid tail or polynucleotide sequences to increase the immunogenic potential of the final molecule is also considered.

- 5 In one aspect, the invention relates to the use of genetically engineered soluble fusion proteins comprising a polypeptide of the present invention, or a fragment thereof, and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclasses (IgG, IgM, IgA, IgE). Preferred as an immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes
10 place at the hinge region. In a particular embodiment, the Fc part can be removed simply by incorporation of a cleavage sequence which can be cleaved with blood clotting factor Xa.

- Examples of fusion protein technology can be found in International Patent Application
15 Nos. WO94/29458 and WO94/22914.

- The proteins may be chemically conjugated, or expressed as recombinant fusion proteins allowing increased levels to be produced in an expression system as compared to non-fused protein. The fusion partner may assist in providing T helper epitopes
20 (immunological fusion partner), preferably T helper epitopes recognised by humans, or assist in expressing the protein (expression enhancer) at higher yields than the native recombinant protein. Preferably the fusion partner will be both an immunological fusion partner and expression enhancing partner.

- 25 Fusion partners include protein D from *Haemophilus influenzae* and the non-structural protein from influenza virus, NS1 (hemagglutinin). Another fusion partner is the protein known as LytA. Preferably the C terminal portion of the molecule is used. LytA is derived from *Streptococcus pneumoniae* which synthesizes an N-acetyl-L-alanine

amidase LytA, (coded by the *lytA* gene {Gene, 43 (1986) page 265-272}) an autolysin that specifically degrades certain bonds in the peptidoglycan backbone. The C-terminal domain of the LytA protein is responsible for the affinity to the choline or to some choline analogues such as DEAE. This property has been exploited for the development of *E.coli* C- LytA expressing plasmids useful for expression of fusion proteins.
Purification of hybrid proteins containing the C- LytA fragment at its amino terminus has been described {Biotechnology: 10, (1992) page 795-798}. It is possible to use the repeat portion of the LytA molecule found in the C terminal end starting at residue 178, for example residues 188 - 305.

10

The present invention also includes variants of the aforementioned polypeptides, that is polypeptides that vary from the referents by conservative amino acid substitutions, whereby a residue is substituted by another with like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr.

15

Polypeptides for use in the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

20

It is most preferred that a polypeptide for use in the invention is derived from *Haemophilus influenzae*, however, it may preferably be obtained from other organisms of the same taxonomic genus. A polypeptide for use in the invention may also be obtained, for example, from organisms of the same taxonomic family or order.

25

Polynucleotides

It is an object of the invention to provide uses for polynucleotides that encode BASB070 polypeptides, particularly polynucleotides that encode the polypeptide herein designated
5 BASB070, for use in or in preparation of the vaccine compositions described herein.

In a particularly preferred embodiment the polynucleotide comprises a region encoding BASB070 polypeptides comprising a sequence set out in SEQ ID NO:1 or 3 which includes a full length gene, or a variant thereof.

10

The BASB070 polynucleotides provided in SEQ ID NO:1 or 3 are the BASB070 polynucleotides from *Haemophilus influenzae* strains Rd KW20 and ntHi 3224.

Using the information provided herein, such as a polynucleotide sequence set out in SEQ ID
15 NO:1 or 3, a polynucleotide of the invention encoding BASB070 polypeptide may be obtained using standard cloning and screening methods, such as those for cloning and sequencing chromosomal DNA fragments from bacteria, followed by obtaining a full length clone. For example, to obtain a polynucleotide sequence for use in the invention, such as a polynucleotide sequence given in SEQ ID NO:1 or 3, typically a library of clones of
20 chromosomal DNA of *Haemophilus influenzae* in *E.coli* or some other suitable host is probed with a radiolabeled oligonucleotide, preferably a 17-mer or longer, derived from a partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using stringent hybridization conditions. By sequencing the individual clones thus identified by hybridization with sequencing primers designed from the
25 original polypeptide or polynucleotide sequence it is then possible to extend the polynucleotide sequence in both directions to determine a full length gene sequence. Conveniently, such sequencing is performed, for example, using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by

Maniatis, T., Fritsch, E.F. and Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989). (see in particular Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70). Direct genomic DNA
5 sequencing may also be performed to obtain a full length gene sequence.

Moreover, the DNA sequence set out in SEQ ID NO:1 or 3 contains an open reading frame encoding a protein having about the number of amino acid residues set forth in SEQ ID NO:2 or 4 with a deduced molecular weight that can be calculated using amino acid residue
10 molecular weight values well known to those skilled in the art.

The polynucleotide of SEQ ID NO:1, between the start codon at nucleotide number 1 and the stop codon which begins at nucleotide number 2740 of SEQ ID NO:1, encodes the polypeptide of SEQ ID NO:2.
15 The polynucleotide of SEQ ID NO:3, between the start codon at nucleotide number 1 and the stop codon which begins at nucleotide number 2755 of SEQ ID NO:3, encodes the polypeptide of SEQ ID NO:4.

In a further aspect, the present invention provides uses for an isolated polynucleotide
20 comprising or consisting of:

- (a) a polynucleotide sequence which has at least 85% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% or exact identity to SEQ ID NO:1 or 3 over the entire length of SEQ ID NO:1 or 3 respectively; or
- 25 (b) a polynucleotide sequence encoding a polypeptide which has at least 85% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% or 100% exact, to the amino acid sequence of SEQ ID NO:2 or 4 over the entire length of SEQ ID NO:2 or 4 respectively.

- A polynucleotide encoding a polypeptide for use in the present invention, including homologues and orthologs from species other than *Haemophilus influenzae*, may be obtained by a process which comprises the steps of screening an appropriate library under
5 stringent hybridization conditions (for example, using a temperature in the range of 45 – 65° C and an SDS concentration from 0.1 – 1%) with a labeled or detectable probe consisting of or comprising the sequence of SEQ ID NO:1 or 3 or a fragment thereof; and isolating a full-length gene and/or genomic clones containing said polynucleotide sequence.
- 10 The invention provides uses for a polynucleotide sequence identical over its entire length to a coding sequence (open reading frame) in SEQ ID NO:1 or 3. Also provided by the invention are uses for a coding sequence for a mature polypeptide or a fragment thereof, by itself as well as a coding sequence for a mature polypeptide or a fragment in reading frame with another coding sequence, such as a sequence encoding a leader or secretory sequence, a
15 pre-, or pro- or prepro-protein sequence. The polynucleotide may also contain at least one non-coding sequence, including for example, but not limited to at least one non-coding 5' and 3' sequence, such as the transcribed but non-translated sequences, termination signals (such as rho-dependent and rho-independent termination signals), ribosome binding sites, Kozak sequences, sequences that stabilize mRNA, introns, and polyadenylation signals.
- 20 The polynucleotide sequence may also comprise additional coding sequence encoding additional amino acids. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain embodiments of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc. Natl. Acad. Sci., USA* 86: 821-824 (1989), or an HA peptide
25 tag (Wilson *et al.*, *Cell* 37: 767 (1984), both of which may be useful in purifying polypeptide sequence fused to them. Polynucleotides for use with the invention also include, but are not limited to, polynucleotides comprising a structural gene and its naturally associated sequences that control gene expression.

The nucleotide sequence encoding BASB070 polypeptide of SEQ ID NO:2 or 4 may be identical to the polypeptide encoding sequence contained in nucleotides 1 to 2739 of SEQ ID NO:1, or the polypeptide encoding sequence contained in nucleotides 1 to 2754 of SEQ ID NO:3, respectively. Alternatively it may be a sequence, which as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:2 or 4.

The term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides that include a sequence encoding a polypeptide of the invention, particularly a bacterial polypeptide and more particularly a polypeptide of the *Haemophilus influenzae* BASB070 having an amino acid sequence set out in SEQ ID NO:2 or 4. The term also encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, polynucleotides interrupted by integrated phage, an integrated insertion sequence, an integrated vector sequence, an integrated transposon sequence, or due to RNA editing or genomic DNA reorganization) together with additional regions, that also may contain coding and/or non-coding sequences.

The invention further relates to variants of the polynucleotides described herein that encode variants of a polypeptide having a deduced amino acid sequence of SEQ ID NO:2 or 4. Fragments of polynucleotides of the invention may be used, for example, to synthesize full-length polynucleotides of the invention.

Further particularly preferred embodiments are polynucleotides encoding BASB070 variants, that have the amino acid sequence of BASB070 polypeptide of SEQ ID NO:2 or 4 in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, modified, deleted and/or added, in any combination. Especially preferred among these are

silent substitutions, additions and deletions, that do not alter the properties and activities of BASB070 polypeptide.

Further preferred for use in the invention are polynucleotides that are at least 85% identical
5 over their entire length to a polynucleotide encoding BASB070 polypeptide having an amino acid sequence set out in SEQ ID NO:2 or 4, and polynucleotides that are complementary to such polynucleotides. Alternatively, most highly preferred are polynucleotides that comprise a region that is at least 90% identical over its entire length to a polynucleotide encoding BASB070 polypeptide and polynucleotides complementary
10 thereto. In this regard, polynucleotides at least 95% identical over their entire length to the same are particularly preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at least 98% and at least 99% are particularly highly preferred, with at least 99% being the more preferred.

15 Preferred embodiments are polynucleotides encoding polypeptides that retain substantially the same biological function or activity as the mature polypeptide encoded by a DNA of SEQ ID NO:1 or 3.

In accordance with certain preferred embodiments of this invention there are provided
20 polynucleotides that hybridize, particularly under stringent conditions, to BASB070 polynucleotide sequences, such as the polynucleotides in SEQ ID NO:1 or 3.

The invention further relates to polynucleotides that hybridize to the polynucleotide sequences provided herein. In this regard, the invention especially relates to polynucleotides
25 that hybridize under stringent conditions to the polynucleotides described herein. As herein used, the terms "stringent conditions" and "stringent hybridization conditions" mean hybridization occurring only if there is at least 95% and preferably at least 97% identity between the sequences. A specific example of stringent hybridization conditions is

overnight incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/ml of denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at about 65°C.

5 Hybridization and wash conditions are well known and exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), particularly Chapter 11 therein. Solution hybridization may also be used with the polynucleotide sequences provided by the invention.

10 A coding region of a BASB070 gene may be isolated by screening using a DNA sequence provided in SEQ ID NO:1 or 3 to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

15

There are several methods available and well known to those skilled in the art to obtain full-length DNAs, or extend short DNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, for example, Frohman, *et al.*, *PNAS USA* 85: 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon™

20 technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon™ technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and an 'adaptor' sequence ligated onto each end. Nucleic acid amplification (PCR) is then carried out to amplify the "missing" 5' end of the DNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using "nested" primers, that is, primers
25 designed to anneal within the amplified product (typically an adaptor specific primer that anneals further 3' in the adaptor sequence and a gene specific primer that anneals further 5' in the selected gene sequence). The products of this reaction can then be analyzed by

DNA sequencing and a full-length DNA constructed either by joining the product directly to the existing DNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the 5' primer.

- 5 The invention also provides uses for polynucleotides that encode a polypeptide that is the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, may allow protein transport, may lengthen or shorten protein
- 10 half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in vivo*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

- A precursor protein, having a mature form of the polypeptide fused to one or more
- 15 prosequences may be an inactive form of the polypeptide. When prosequences are removed such inactive precursors generally are activated. Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

- In accordance with one particular aspect of the invention, there is provided the use of a
- 20 polynucleotide as described herein for therapeutic or prophylactic purposes, in particular genetic immunization. This is described in more detail later on in the section headed "Vaccines".

- The use of a polynucleotide of the invention in genetic immunization will preferably
- 25 employ a suitable delivery method such as direct injection of plasmid DNA into muscles (Wolff *et al.*, *Hum Mol Genet* (1992) 1: 363, Manthorpe *et al.*, *Hum. Gene Ther.* (1983) 4: 419), delivery of DNA complexed with specific protein carriers (Wu *et al.*, *J Biol Chem.* (1989) 264: 16985), coprecipitation of DNA with calcium phosphate (Benvenisty &

Reshef, *PNAS USA*, (1986) 83: 9551), encapsulation of DNA in various forms of liposomes (Kaneda *et al.*, *Science* (1989) 243: 375), particle bombardment (Tang *et al.*, *Nature* (1992) 356:152, Eisenbraun *et al.*, *DNA Cell Biol* (1993) 12: 791) and *in vivo* infection using cloned retroviral vectors (Seeger *et al.*, *PNAS USA* (1984) 81: 5849).

5

Vectors, Host Cells, Expression Systems

The invention relates to vectors that comprise a polynucleotide or polynucleotides of the invention, host cells that are genetically engineered with vectors of the invention and the
10 production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the invention.

Recombinant polypeptides for use in the present invention may be prepared using
15 genetically engineered host cells comprising expression systems by processes well known in the art. Accordingly, in a further aspect, the present invention relates to expression systems that comprise a polynucleotide or polynucleotides of the present invention, to host cells which are genetically engineered with such expression systems, and to the production of polypeptides of the invention by recombinant techniques.

20

For recombinant production of the polypeptides of the invention, host cells can be genetically engineered to incorporate expression systems or portions thereof or polynucleotides of the invention. Introduction of a polynucleotide into the host cell can be effected by methods described in many standard laboratory manuals, such as Davis, *et al.*, *BASIC METHODS IN MOLECULAR BIOLOGY*, (1986) and Sambrook, *et al.*,
25 *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), such as, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic

lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction and infection.

Representative examples of appropriate hosts include bacterial cells, such as cells of
5 *streptococci*, *staphylococci*, *enterococci*, *E. coli*, *streptomyces*, *cyanobacteria*, *Bacillus subtilis*, *Neisseria*, *Moraxella* and *Haemophilus influenzae*; fungal cells, such as cells of a yeast, *Kluveromyces*, *Saccharomyces*, a basidiomycete, *Candida albicans* and *Aspergillus*; insect cells such as cells of *Drosophila* S2 and *Spodoptera* Sf9; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293, CV-1 and Bowes melanoma cells; and plant cells, such
10 as cells of a gymnosperm or angiosperm.

A great variety of expression systems can be used to produce the polypeptides of the invention. Such vectors include, among others, chromosomal-, episomal- and virus-derived vectors, for example, vectors derived from bacterial plasmids, from bacteriophage, from
15 transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses, picornaviruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The
20 expression system constructs may contain control regions that regulate as well as engender expression. Generally, any system or vector suitable to maintain, propagate or express polynucleotides and/or to express a polypeptide in a host may be used for expression in this regard. The appropriate DNA sequence may be inserted into the expression system by any
25 of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL*, (*supra*).

Polypeptides of the invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid

extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxyapatite chromatography, and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well-known techniques for refolding protein
5 may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

The expression system may also be a recombinant live microorganism, such as a virus or bacterium. The gene of interest can be inserted into the genome of a live recombinant
10 virus or bacterium. Inoculation and *in vivo* infection with this live vector will lead to *in vivo* expression of the antigen and induction of immune responses. Viruses and bacteria used for this purpose are for instance: poxviruses (e.g; vaccinia, fowlpox, canarypox), alphaviruses (Sindbis virus, Semliki Forest Virus, Venezuelan Equine Encephalitis Virus), adenoviruses, adeno-associated virus, picornaviruses (poliovirus, rhinovirus),
15 herpesviruses (varicella zoster virus, etc), *Listeria*, *Salmonella*, *Shigella*, BCG. These viruses and bacteria can be virulent, or attenuated in various ways in order to obtain live vaccines. Such live vaccines also form part of the invention.

Diagnostic, Prognostic, Serotyping and Mutation Assays

20 This invention is also related to the use of BASB070 polynucleotides and polypeptides of the invention for use as diagnostic reagents. Detection of BASB070 polynucleotides and/or polypeptides in a eukaryote, particularly a mammal, and especially a human, will provide a diagnostic method for diagnosis of disease, staging of disease or response of an infectious
25 organism to drugs. Eukaryotes, particularly mammals, and especially humans, particularly those infected or suspected to be infected with an organism comprising the BASB070 gene or protein, may be detected at the nucleic acid or amino acid level by a variety of well known techniques as well as by methods provided herein.

- Polypeptides and polynucleotides for prognosis, diagnosis or other analysis may be obtained from a putatively infected and/or infected individual's bodily materials. Polynucleotides from any of these sources, particularly DNA or RNA, may be used directly for detection or may be amplified enzymatically by using PCR or any other amplification technique prior to analysis. RNA, particularly mRNA, cDNA and genomic DNA may also be used in the same ways. Using amplification, characterization of the species and strain of infectious or resident organism present in an individual, may be made by an analysis of the genotype of a selected polynucleotide of the organism. Deletions and insertions can be detected by a change in size of the amplified product in comparison to a genotype of a reference sequence selected from a related organism, preferably a different species of the same genus or a different strain of the same species. Point mutations can be identified by hybridizing amplified DNA to labeled BASB070 polynucleotide sequences. Perfectly or significantly matched sequences can be distinguished from imperfectly or more significantly mismatched duplexes by DNase or RNase digestion, for DNA or RNA respectively, or by detecting differences in melting temperatures or renaturation kinetics. Polynucleotide sequence differences may also be detected by alterations in the electrophoretic mobility of polynucleotide fragments in gels as compared to a reference sequence. This may be carried out with or without denaturing agents. Polynucleotide differences may also be detected by direct DNA or RNA sequencing. See, for example, Myers *et al.*, *Science*, 230: 1242 (1985). Sequence changes at specific locations also may be revealed by nuclease protection assays, such as RNase, V1 and S1 protection assay or a chemical cleavage method. See, for example, Cotton *et al.*, *Proc. Natl. Acad. Sci., USA*, 85: 4397-4401 (1985).
- In another embodiment, an array of oligonucleotides probes comprising BASB070 nucleotide sequence or fragments thereof can be constructed to conduct efficient screening of, for example, genetic mutations, serotype, taxonomic classification or identification. Array technology methods are well known and have general applicability and can be used to

address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see, for example, Chee *et al.*, *Science*, 274: 610 (1996)).

Thus in another aspect, the present invention relates to a diagnostic kit which comprises:

- 5 (a) a polynucleotide of the present invention, preferably the nucleotide sequence of SEQ ID NO:1 or 3, or a fragment thereof ; (b) a nucleotide sequence complementary to that of (a); (c) a polypeptide of the present invention, preferably the polypeptide of SEQ ID NO:2 or 4 or a fragment thereof; or (d) an antibody to a polypeptide of the present invention, preferably to the polypeptide of SEQ ID NO:2 or 4.

10

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component. Such a kit will be of use in diagnosing a disease or susceptibility to a Disease, among others.

- 15 This invention also relates to the use of polynucleotides of the present invention as diagnostic reagents. Detection of a mutated form of a polynucleotide of the invention, preferably SEQ ID NO:1 or 3, which is associated with a disease or pathogenicity will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, a prognosis of a course of disease, a determination of a stage of disease, or a susceptibility to a disease,
- 20 which results from under-expression, over-expression or altered expression of the polynucleotide. Organisms, particularly infectious organisms, carrying mutations in such polynucleotide may be detected at the polynucleotide or polypeptide level by a variety of techniques, such as those described elsewhere herein.

- 25 The nucleotide sequences of the present invention are also valuable for organism chromosome identification. The sequence is specifically targeted to, and can hybridize with, a particular location on an organism's chromosome, particularly to a *Haemophilus influenzae* chromosome. The mapping of relevant sequences to chromosomes according to

the present invention may be an important step in correlating those sequences with pathogenic potential and/or an ecological niche of an organism and/or drug resistance of an organism, as well as the essentiality of the gene to the organism. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data may be found on-line in a sequence database. The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through known genetic methods, for example, through linkage analysis (coinheritance of physically adjacent genes) or mating studies, such as by conjugation.

The differences in a polynucleotide and/or polypeptide sequence between organisms possessing a first phenotype and organisms possessing a different, second different phenotype can also be determined. If a mutation is observed in some or all organisms possessing the first phenotype but not in any organisms possessing the second phenotype, then the mutation is likely to be the causative agent of the first phenotype.

Cells from an organism carrying mutations or polymorphisms (allelic variations) in a polynucleotide and/or polypeptide of the invention may also be detected at the polynucleotide or polypeptide level by a variety of techniques, to allow for serotyping, for example. For example, RT-PCR can be used to detect mutations in the RNA. It is particularly preferred to use RT-PCR in conjunction with automated detection systems, such as, for example, GeneScan. RNA, cDNA or genomic DNA may also be used for the same purpose, PCR. As an example, PCR primers complementary to a polynucleotide encoding BASB070 polypeptide can be used to identify and analyze mutations.

The invention further provides primers for, among other things, amplifying BASB070 DNA and/or RNA isolated from a sample derived from an individual, such as a bodily material. The primers may be used to amplify a polynucleotide isolated from an infected individual,

such that the polynucleotide may then be subject to various techniques for elucidation of the polynucleotide sequence. In this way, mutations in the polynucleotide sequence may be detected and used to diagnose and/or give a prognosis for the infection or its stage or course, or to serotype and/or classify the infectious agent.

5

The invention further provides a process for diagnosing, disease, preferably bacterial infections, more preferably infections caused by *Haemophilus influenzae*, comprising determining from a sample derived from an individual, such as a bodily material, an increased level of expression of polynucleotide having a sequence of Table 1 [SEQ ID
10 NO:1 or 3]. Increased or decreased expression of a BASB070 polynucleotide can be measured using any one of the methods well known in the art for the quantitation of polynucleotides, such as, for example, amplification, PCR, RT-PCR, RNase protection, Northern blotting, spectrometry and other hybridization methods.

15 In addition, a diagnostic assay in accordance with the invention for detecting over- or under-expression of BASB070 polypeptide compared to normal control tissue samples may be used to detect the presence of an infection, for example. Assay techniques that can be used to determine levels of a BASB070 polypeptide, in a sample derived from a host, such as a bodily material, are well known to those of skill in the art. Such assay methods include
20 radioimmunoassays, competitive-binding assays, Western Blot analysis, antibody sandwich assays, antibody detection and ELISA assays.

The polynucleotides of the invention may be used as components of polynucleotide arrays, preferably high-density arrays or grids. These high-density arrays are
25 particularly useful for diagnostic and prognostic purposes. For example, a set of spots each comprising a different gene, and further comprising a polynucleotide or polynucleotides of the invention, may be used for probing, such as using hybridization or nucleic acid amplification, using probes obtained or derived from a bodily sample, to

determine the presence of a particular polynucleotide sequence or related sequence in an individual. Such a presence may indicate the presence of a pathogen, particularly *Haemophilus influenzae*, and may be useful in diagnosing and/or giving a prognosis for disease or a course of disease. A grid comprising a number of variants of the polynucleotide sequence of SEQ ID NO: 1 is preferred. Also preferred is a grid comprising a number of variants of a polynucleotide sequence encoding the polypeptide sequence of SEQ ID NO: 2.

Antibodies

10

The polypeptides and polynucleotides of the invention or variants thereof, or cells expressing the same can be used as immunogens to produce antibodies immunospecific for such polypeptides or polynucleotides respectively.

15 In certain preferred embodiments of the invention there are provided antibodies against BASB070 polypeptides or polynucleotides.

Antibodies generated against the polypeptides or polynucleotides of the invention can be obtained by administering the polypeptides and/or polynucleotides of the invention, or epitope-bearing fragments of either or both, analogues of either or both, or cells expressing either or both, to an animal, preferably a nonhuman, using routine protocols. For preparation of monoclonal antibodies, any technique known in the art that provides antibodies produced by continuous cell line cultures can be used. Examples include various techniques, such as those in Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, pg. 77-96 in *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc. (1985).

Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to polypeptides or polynucleotides of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies immunospecific to the polypeptides or polynucleotides of the invention.

Alternatively, phage display technology may be utilized to select genes for antibodies with binding activities towards a polypeptide of the invention either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing anti-BASB070 or from naive libraries (McCafferty, *et al.*, (1990), *Nature* 348, 552-554; Marks, *et al.*, (1992) *Biotechnology* 10, 779-783). The affinity of these antibodies can also be improved by, for example, chain shuffling (Clackson *et al.*, (1991) *Nature* 352: 628).

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptides or polynucleotides of the invention to purify the polypeptides or polynucleotides by, for example, affinity chromatography.

Thus, among others, antibodies against BASB070-polypeptide or BASB070-polynucleotide may be employed to treat infections, particularly bacterial infections.

Polypeptide variants include antigenically, epitopically or immunologically equivalent variants form a particular aspect of this invention.

Preferably, the antibody or variant thereof is modified to make it less immunogenic in the individual. For example, if the individual is human the antibody may most preferably be "humanized," where the complementarity determining region or regions of the hybridoma-derived antibody has been transplanted into a human monoclonal antibody, for

example as described in Jones *et al.* (1986), *Nature* 321, 522-525 or Tempest *et al.*, (1991) *Biotechnology* 9, 266-273.

In a further aspect, the present invention relates to genetically engineered soluble fusion
5 proteins comprising a polypeptide of the present invention, or a fragment thereof, and
various portions of the constant regions of heavy or light chains of immunoglobulins of
various subclasses (IgG, IgM, IgA, IgE). Preferred as an immunoglobulin is the constant
part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the
hinge region. In a particular embodiment, the Fc part can be removed simply by
10 incorporation of a cleavage sequence which can be cleaved with blood clotting factor Xa.
Furthermore, this invention relates to processes for the preparation of these fusion
proteins by genetic engineering, and to the use thereof for drug screening, diagnosis and
therapy. A further aspect of the invention also relates to polynucleotides encoding such
fusion proteins. Examples of fusion protein technology can be found in International
15 Patent Application Nos. WO94/29458 and WO94/22914.

Mimotopes

In a further aspect, the present invention relates to mimotopes of the polypeptide of the
20 invention. A mimotope is generally a peptide sequence, sufficiently similar to the
native peptide (sequentially or structurally), which is capable of binding to the binding
site of the native peptide. Thus where an antibody-binding peptide is concerned, a
mimotope is capable of being recognised by antibodies which recognise the native
peptide; or is capable of raising antibodies which recognise the native peptide,
25 optionally when coupled to a suitable carrier. In the case of T cell recognition, a
mimotope is capable of being recognised by the same T cells that recognise the native
peptide; or is capable of generating a T cell response which recognises the native
peptide.

Peptide mimotopes may be designed for a particular purpose by addition, deletion or substitution of elected amino acids. Thus, the peptides may be modified for the purposes of ease of conjugation to a protein carrier. For example, it may be desirable for some chemical conjugation methods to include a terminal cysteine. In addition it may be desirable for peptides conjugated to a protein carrier to include a hydrophobic terminus distal from the conjugated terminus of the peptide, such that the free unconjugated end of the peptide remains associated with the surface of the carrier protein. Thereby presenting the peptide in a conformation which most closely resembles that of the peptide as found in the context of the whole native molecule. For example, the peptides may be altered to have an N-terminal cysteine and a C-terminal hydrophobic amidated tail. Alternatively, the addition or substitution of a D-stereoisomer form of one or more of the amino acids may be performed to create a beneficial derivative, for example to enhance stability of the peptide and/or to increase the affinity of the peptide for a particular ligand.

Mimotopes may also be retro sequences of the natural peptide sequences, in that the sequence orientation is reversed; or alternatively the sequences may be entirely or at least in part comprised of D-stereoisomer amino acids (inverso sequences). Also, the peptide sequences may be retro-inverso in character, in that the sequence orientation is reversed and the amino acids are of the D-stereoisomer form. Retro, inverso and retro-inverso peptides are described in WO95/24916 and WO94/05311.

Alternatively, peptide mimotopes may be identified using antibodies which are capable themselves of binding to the polypeptides of the present invention using techniques such as phage display technology (EP 0 552 267 B1). This technique, generates a large number of peptide sequences which mimic the structure of the native peptides and are, therefore,

capable of binding to anti-native peptide antibodies, but may not necessarily themselves share significant sequence homology to the native polypeptide.

Vaccines

5

One particularly important aspect of the invention relates to a method for inducing an immunological response in an individual, particularly a mammal, preferably humans, which comprises inoculating the individual with a BASB070 polynucleotide and/or polypeptide, or a fragment, or a mimotope, or a variant thereof, adequate to produce antibody and/or T cell immune response to protect said individual from infection, particularly bacterial infection and most particularly *Haemophilus influenzae* infection. Also provided are methods whereby such immunological response slows bacterial replication.

15

Yet another aspect of the invention relates to a method of inducing an immunological response in an individual which comprises delivering to such individual a nucleic acid vector, sequence or ribozyme to direct expression of BASB070 polynucleotide and/or polypeptide, or a fragment, or a mimotope, or a variant thereof, for expressing BASB070 polynucleotide and/or polypeptide, or a fragment, or a mimotope, or a variant thereof *in vivo* in order to induce an immunological response, such as, to produce antibody and/or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said individual, preferably a human, from disease, whether that disease is already established within the individual or not. One example of administering the gene is by accelerating it into the desired cells as a coating on particles or otherwise. Such nucleic acid vectors may comprise DNA, RNA, a ribozyme, a modified nucleic acid, a DNA/RNA hybrid, a DNA-protein complex or an RNA-protein complex.

25

A further aspect of the invention relates to an immunological composition that when introduced into an individual, preferably a human, capable of having induced within it an immunological response, induces an immunological response in such individual to a BASB070 polynucleotide and/or polypeptide encoded therefrom, wherein the composition
5 comprises a recombinant BASB070 polynucleotide and/or polypeptide encoded therefrom, or a fragment, or a mimotope, or a variant thereof, and/or comprises DNA and/or RNA which encodes and expresses an antigen of said BASB070 polynucleotide, polypeptide encoded therefrom, or other polypeptide of the invention, such as a fragment or a mimotope or a variant. The immunological response may be used therapeutically or
10 prophylactically and may take the form of antibody immunity and/or cellular immunity, such as cellular immunity arising from CTL or CD4+ T cells.

A BASB070 polypeptide or a fragment thereof may be fused with a co-protein or chemical moiety which may or may not by itself produce antibodies or induce a T cell
15 response, but which is capable of stabilizing the first protein and producing a fused or modified protein which will have antigenic and/or immunogenic properties, and preferably protective properties. Thus fused recombinant protein, preferably further comprises an antigenic co-protein, such as lipoprotein D from *Haemophilus influenzae*, Glutathione-S-transferase (GST) or beta-galactosidase, or any other relatively large co-
20 protein which solubilizes the protein and facilitates production and purification thereof. Moreover, the co-protein may act as an adjuvant in the sense of providing a generalized stimulation of the immune system of the organism receiving the protein. The co-protein may be attached to either the amino- or carboxy-terminus of the first protein.

25 In a vaccine composition according to the invention, a BASB070 polynucleotide and/or polypeptide, or a fragment, or a mimotope, or a variant thereof may be present in or encoded by a vector, such as the live recombinant vectors described above for example live bacterial vectors.

Also suitable are non-live vectors for the BASB070 polypeptide, for example bacterial outer-membrane vesicles or "blebs". OM blebs are derived from the outer membrane of the two-layer membrane of Gram-negative bacteria and have been documented in many Gram-negative bacteria (Zhou, L *et al.* 1998. *FEMS Microbiol. Lett.* 163:223-228) including *C. trachomatis* and *C. psittaci*. A non-exhaustive list of bacterial pathogens reported to produce blebs also includes: *Bordetella pertussis*, *Borrelia burgdorferi*, *Brucella melitensis*, *Brucella ovis*, *Escherichia coli*, *Haemophilus influenza*, *Legionella pneumophila*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Pseudomonas aeruginosa* and *Yersinia enterocolitica*.

Blebs have the advantage of providing outer-membrane proteins in their native conformation and are thus particularly useful for vaccines. Blebs can also be improved for vaccine use by engineering the bacterium so as to modify the expression of one or more molecules at the outer membrane. Thus for example the expression of a desired immunogenic protein at the outer membrane, such as the BASB070 polypeptide, can be introduced or upregulated (e.g. by altering the promoter). Instead or in addition, the expression of outer-membrane molecules which are either not relevant (e.g. unprotective antigens or immunodominant but variable proteins) or detrimental (e.g. toxic molecules such as LPS, or potential inducers of an autoimmune response) can be downregulated. These approaches are discussed in more detail below.

The non-coding flanking regions of the BASB070 gene contain regulatory elements important in the expression of the gene. This regulation takes place both at the transcriptional and translational level. The sequence of these regions, either upstream or downstream of the open reading frame of the gene, can be obtained by DNA sequencing. This sequence information allows the determination of potential regulatory motifs such as the different promoter elements, terminator sequences, inducible sequence elements,

repressors, elements responsible for phase variation, the shine-dalgarno sequence, regions with potential secondary structure involved in regulation, as well as other types of regulatory motifs or sequences.

5 This sequence information allows the modulation of the natural expression of the BASB070 gene. The upregulation of the gene expression may be accomplished by altering the promoter, the shine-dalgarno sequence, potential repressor or operator elements, or any other elements involved. Likewise, downregulation of expression can be achieved by similar types of modification. Alternatively, by changing phase variation
10 sequences, the expression of the gene can be put under phase variation control, or it may be uncoupled from this regulation. In another approach, the expression of the gene can be put under the control of one or more inducible elements allowing regulated expression. Examples of such regulation include, but are not limited to, induction by temperature shift, addition of inductor substrates like selected carbohydrates or their derivatives, trace
15 elements, vitamins, co-factors, metal ions, etc.

Such modifications as described above can be introduced by several different means. The modification of sequences involved in gene expression can be carried out *in vivo* by random mutagenesis followed by selection for the desired phenotype. Another approach
20 consists in isolating the region of interest and modifying it by random mutagenesis, or site-directed replacement, insertion or deletion mutagenesis. The modified region can then be reintroduced into the bacterial genome by homologous recombination, and the effect on gene expression can be assessed. In another approach, the sequence knowledge of the region of interest can be used to replace or delete all or part of the natural regulatory
25 sequences. In this case, the regulatory region targeted is isolated and modified so as to contain the regulatory elements from another gene, a combination of regulatory elements from different genes, a synthetic regulatory region, or any other regulatory region, or to delete selected parts of the wild-type regulatory sequences. These modified sequences can

then be reintroduced into the bacterium via homologous recombination into the genome. A non-exhaustive list of preferred promoters that could be used for up-regulation of gene expression includes the promoters *porA*, *porB*, *lbpB*, *tbpB*, *p110*, *lst*, *hpuAB* from *N. meningitidis* or *N. gonorrhoea*; *ompCD*, *copB*, *lbpB*, *ompE*, *UspA1*, *UspA2*, *TbpB* from *M. Catarrhalis*; *p1*, *p2*, *p4*, *p5*, *p6*, *lpD*, *tbpB*, *D15*, *Hia*, *Hmw1*, *Hmw2* from *H. influenzae*.

In one example, the expression of the gene can be modulated by exchanging its promoter with a stronger promoter (through isolating the upstream sequence of the gene, in vitro modification of this sequence, and reintroduction into the genome by homologous recombination). Upregulated expression can be obtained in both the bacterium as well as in the outer membrane vesicles shed (or made) from the bacterium.

In other examples, the described approaches can be used to generate recombinant bacterial strains with improved characteristics for vaccine applications. These can be, but are not limited to, attenuated strains, strains with increased expression of selected antigens, strains with knock-outs (or decreased expression) of genes interfering with the immune response, strains with modulated expression of immunodominant proteins, strains with modulated shedding of outer-membrane vesicles.

Thus, also provided by the invention is a modified upstream region of the BASB070 gene, which modified upstream region contains a heterologous regulatory element which alters the expression level of the BASB070 protein located at the outer membrane. The upstream region according to this aspect of the invention includes the sequence upstream of the BASB070 gene. The upstream region starts immediately upstream of the BASB070 gene and continues usually to a position no more than about 1000 bp upstream of the gene from the ATG start codon. In the case of a gene located in a polycistronic sequence (operon) the upstream region can start immediately preceding the gene of interest, or

preceding the first gene in the operon. Preferably, a modified upstream region according to this aspect of the invention contains a heterologous promotor at a position between 500 and 700 bp upstream of the ATG.

- 5 Thus, the invention provides the BASB070 polypeptide, in a modified bacterial Bleb. The invention further provides modified host cells capable of producing the non-live membrane-based bleb vectors. The invention further provides vectors comprising the BASB070 gene having a modified upstream region containing a heterologous regulatory element.
- 10 Further provided by the invention are processes to prepare the host cells and bacterial blebs according to the invention.

Vaccine antigens may be provided in a variety of other forms known in the art, depending on the properties of the protein. Lipoproteins for example, because of the hydrophobicity
15 of the lipids added to their N-terminus, are able to aggregate and to form micelles. The particulate nature of these structures can enhance the immunogenicity of the lipoprotein, as compared to the unlipidated version of the protein. The size of the micelles may also have an impact on the immunogenicity of the lipoprotein and this can be modified for example by adjusting the extraction procedure.

20 Also provided by this invention are compositions, particularly vaccine compositions, and methods comprising the polypeptides and/or polynucleotides of the invention and immunostimulatory DNA sequences, such as those described in Sato, Y. et al. Science 273: 352 (1996).

25 Also, provided by this invention are methods using the described polynucleotide or particular fragments thereof, which have been shown to encode non-variable regions of bacterial cell surface proteins, in polynucleotide constructs used in such genetic

immunization experiments in animal models of infection with *Haemophilus influenzae*. Such experiments will be particularly useful for identifying protein epitopes able to provoke a prophylactic or therapeutic immune response. It is believed that this approach will allow for the subsequent preparation of monoclonal antibodies of particular value,
5 derived from the requisite organ of the animal successfully resisting or clearing infection, for the development of prophylactic agents or therapeutic treatments of bacterial infection, particularly *Haemophilus influenzae* infection, in mammals, particularly humans.

The invention also includes a vaccine formulation which comprises an immunogenic
10 recombinant polypeptide and/or polynucleotide of the invention together with a suitable carrier, such as any pharmaceutically acceptable carrier. Since the polypeptides and polynucleotides may be broken down in the stomach, each could be administered via a mucosal surface such as intranasally, or administered parenterally, including, for example, administration that is subcutaneous, intramuscular, intravenous, or intradermal.
15 Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostatic compounds and solutes which render the formulation isotonic with the bodily fluid, preferably the blood, of the individual; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be
20 presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use.

The vaccine formulation of the invention may also include adjuvant systems for
25 enhancing the immunogenicity of the formulation.

An immune response may be broadly distinguished into two extreme categories, being a humoral or cell mediated immune responses (traditionally characterised by antibody and

cellular effector mechanisms of protection respectively). These categories of response have been termed TH1-type responses (cell-mediated response), and TH2-type immune responses (humoral response).

- 5 Extreme TH1-type immune responses may be characterised by the generation of antigen specific, haplotype restricted cytotoxic T lymphocytes, and natural killer cell responses. In mice TH1-type responses are often characterised by the generation of antibodies of the IgG2a subtype, whilst in the human these correspond to IgG1 type antibodies. TH2-type immune responses are characterised by the generation of a broad range of
- 10 immunoglobulin isotypes including in mice IgG1, IgA, and IgM.

- It can be considered that the driving force behind the development of these two types of immune responses are cytokines. High levels of TH1-type cytokines tend to favour the induction of cell mediated immune responses to the given antigen, whilst high levels of
- 15 TH2-type cytokines tend to favour the induction of humoral immune responses to the antigen.

- The distinction of TH1 and TH2-type immune responses is not absolute. In reality an individual will support an immune response which is described as being predominantly
- 20 TH1 or predominantly TH2. However, it is often convenient to consider the families of cytokines in terms of that described in murine CD4 T cell clones by Mosmann and Coffman (*Mosmann, T.R. and Coffman, R.L. (1989) TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties. Annual Review of Immunology, 7, p145-173*). Traditionally, TH1-type responses are associated with
- 25 the production of the IFN- γ and IL-2 cytokines by T-lymphocytes. Other cytokines often directly associated with the induction of TH1-type immune responses are not produced by T-cells, such as IL-12. In contrast, TH2-type responses are associated with the secretion of IL-4, IL-5, IL-6 and IL-13.

It is known that certain vaccine adjuvants are particularly suited to the stimulation of either TH1 or TH2 - type cytokine responses. Traditionally the best indicators of the TH1:TH2 balance of the immune response after a vaccination or infection includes
5 direct measurement of the production of TH1 or TH2 cytokines by T lymphocytes *in vitro* after restimulation with antigen, and/or (in the murine system) the measurement of the IgG1:IgG2a ratio of antigen specific antibody responses.

Thus, a TH1-type adjuvant is one which preferentially stimulates isolated T-cell
10 populations to produce a high ratio of TH1-type cytokines when re-stimulated with antigen *in vitro*, and promotes development of both CD8+ cytotoxic T lymphocytes and antigen specific immunoglobulin responses associated with TH1-type isotype. Adjuvants which are capable of preferential stimulation of the TH1 cell response are described in International Patent Application No. WO 94/00153 and WO 95/17209.

15 3 De-O-acylated monophosphoryl lipid A (3D-MPL) is one such adjuvant. This is known from GB 2220211 (Ribi). Chemically it is a mixture of 3 De-O-acylated monophosphoryl lipid A with 4, 5 or 6 acylated chains and is manufactured by Ribi Immunochem, Montana. A preferred form of 3 De-O-acylated monophosphoryl lipid
20 A is disclosed in European Patent 0 689 454 B1 (SmithKline Beecham Biologicals SA).

Preferably, the particles of 3D-MPL are small enough to be sterile filtered through a 0.22micron membrane (European Patent number 0 689 454).

3D-MPL will be present in the range of 10µg - 100µg preferably 25-50µg per dose
25 wherein the antigen will typically be present in a range 2-50µg per dose.

Another preferred adjuvant comprises QS21, an HPLC purified non-toxic fraction derived from the bark of *Quillaja Saponaria* Molina. Optionally this may be admixed

with 3 De-O-acylated monophosphoryl lipid A (3D-MPL), optionally together with a carrier.

The method of production of QS21 is disclosed in US patent No. 5,057,540.

5

Non-reactogenic adjuvant formulations containing QS21 have been described previously (WO 96/33739). Such formulations comprising QS21 and cholesterol have been shown to be successful TH1 stimulating adjuvants when formulated together with an antigen.

10

Further adjuvants which are preferential stimulators of TH1 cell responses include immunomodulatory oligonucleotides, for example unmethylated CpG sequences as disclosed in WO 96/02555.

15

Combinations of different TH1 stimulating adjuvants, such as those mentioned hereinabove, are also contemplated as providing an adjuvant which is a preferential stimulator of TH1 cell response. For example, QS21 can be formulated together with 3D-MPL. The ratio of QS21 : 3D-MPL will typically be in the order of 1 : 10 to 10 : 1; preferably 1:5 to 5 : 1 and often substantially 1 : 1. The preferred range for optimal synergy is 2.5 : 1 to 1 : 1 3D-MPL: QS21.

20

Preferably a carrier which enhances immunogenicity is also present in the vaccine composition according to the invention. Such a carrier may be an oil in water emulsion, or an aluminium salt, such as aluminium phosphate or aluminium hydroxide.

25

A preferred oil-in-water emulsion comprises a metabolisable oil, such as squalene, alpha tocopherol and Tween 80. In a particularly preferred aspect the antigens in the vaccine composition according to the invention are combined with QS21 and 3D-MPL in such

an emulsion. Additionally the oil in water emulsion may contain span 85 and/or lecithin and/or tricaprylin.

Typically for human administration QS21 and 3D-MPL will be present in a vaccine in the range of 1µg - 200µg, such as 10-100µg, preferably 10µg - 50µg per dose.

Typically the oil in water will comprise from 2 to 10% squalene, from 2 to 10% alpha tocopherol and from 0.3 to 3% tween 80. Preferably the ratio of squalene: alpha tocopherol is equal to or less than 1 as this provides a more stable emulsion. Span 85 may also be present at a level of 1%. In some cases it may be advantageous that the vaccines of the present invention will further contain a stabiliser.

Non-toxic oil in water emulsions preferably contain a non-toxic oil, e.g. squalane or squalene, an emulsifier, e.g. Tween 80, in an aqueous carrier. The aqueous carrier may be, for example, phosphate buffered saline.

A particularly potent adjuvant formulation involving QS21, 3D-MPL and tocopherol in an oil in water emulsion is described in WO 95/17210.

The present invention also provides a polyvalent vaccine composition comprising a vaccine formulation of the invention in combination with other antigens, in particular antigens useful for treating other bacterial or viral diseases, cancers, autoimmune diseases and related conditions. Such a polyvalent vaccine composition may include a TH-1 inducing adjuvant as hereinbefore described.

While the invention has been described with reference to certain BASB070 polypeptides and polynucleotides, it is to be understood that this covers fragments of the naturally occurring polypeptides and polynucleotides, and similar polypeptides and polynucleotides

with additions, deletions or substitutions which do not substantially affect the immunogenic properties of the recombinant polypeptides or polynucleotides.

Compositions, kits and administration

5

In a further aspect of the invention there are provided compositions comprising a BASB070 polynucleotide and/or BASB070 polypeptide for administration to a cell or to a multicellular organism.

10

The invention also relates to compositions comprising a polynucleotide and/or a polypeptides discussed herein or their agonists or antagonists. The polypeptides and polynucleotides of the invention may be employed in combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for administration to an individual. Such compositions comprise, for
15 instance, a media additive or a therapeutically effective amount of a polypeptide and/or polynucleotide of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof. The formulation should suit the mode of administration. The invention further relates to diagnostic and pharmaceutical packs and kits comprising
20 one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

25

Polypeptides, polynucleotides and other compounds of the invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous,

intraperitoneal, intramuscular, subcutaneous, intranasal intradermal or transdermal routes among others.

5 In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

10 In a further aspect, the present invention provides for pharmaceutical compositions comprising a therapeutically effective amount of a polypeptide and/or polynucleotide, such as the soluble form of a polypeptide and/or polynucleotide of the present invention, agonist or antagonist peptide or small molecule compound, in combination with a pharmaceutically acceptable carrier or excipient. Such carriers include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one 15 or more of the ingredients of the aforementioned compositions of the invention. Polypeptides, polynucleotides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

20 The composition will be adapted to the route of administration, for instance by a systemic or an oral route. Preferred forms of systemic administration include injection, typically by intramuscular or subcutaneous injection. Other injection routes, such as intradermal, intraperitoneal, or intravenous can be used. Alternative means for systemic administration include transmucosal and transdermal administration using penetrants such as bile salts or fusidic acids or other detergents. In addition, if a polypeptide or other compounds of the 25 present invention can be formulated in an enteric or an encapsulated formulation, oral administration may also be possible. Administration of these compounds may also be topical and/or localized, in the form of salves, pastes, gels, solutions, powders and the like.

For administration to mammals, and particularly humans, it is expected that the dosage level of the active agent will be from 0.01 $\mu\text{g/kg}$ to 10 $\mu\text{g/kg}$, typically around 1 $\mu\text{g/kg}$. The physician in any event will determine the actual dosage which will be most suitable for an individual and will vary with the age, weight and response of the particular individual. The above dosages are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited, and such are within the scope of this invention.

The dosage range required depends on the choice of peptide, the route of administration, the nature of the formulation, the nature of the subject's condition, and the judgment of the attending practitioner. Suitable dosages, however, are in the range of 0.1-100 $\mu\text{g/kg}$ of subject.

A vaccine composition is conveniently in injectable form. Conventional adjuvants may be employed to enhance the immune response. A suitable unit dose for vaccination is 0.5-5 microgram/kg of antigen, and such dose is preferably administered 1-3 times and with an interval of 1-3 weeks. With the indicated dose range, no adverse toxicological effects will be observed with the compounds of the invention which would preclude their administration to suitable individuals.

Wide variations in the required dosage, however, are to be expected in view of the variety of compounds available and the differing efficiencies of various routes of administration. For example, oral administration would be expected to require higher dosages than administration by injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, as is well understood in the art.

Sequence Databases, Sequences in a Tangible Medium, and Algorithms

Polynucleotide and polypeptide sequences form a valuable information resource with which to determine their 2- and 3-dimensional structures as well as to identify further sequences of similar homology. These approaches are most easily facilitated by storing the sequence in a computer readable medium and then using the stored data in a known macromolecular structure program or to search a sequence database using well known searching tools, such as the GCG program package.

Also provided by the invention are methods for the analysis of character sequences or strings, particularly genetic sequences or encoded protein sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, DNA, RNA and protein structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, codon usage analysis, nucleic acid base trimming, and sequencing chromatogram peak analysis.

A computer based method is provided for performing homology identification. This method comprises the steps of: providing a first polynucleotide sequence comprising the sequence of a polynucleotide of the invention in a computer readable medium; and comparing said first polynucleotide sequence to at least one second polynucleotide or polypeptide sequence to identify homology.

A computer based method is also provided for performing homology identification, said method comprising the steps of: providing a first polypeptide sequence comprising the sequence of a polypeptide of the invention in a computer readable medium; and comparing said first polypeptide sequence to at least one second polynucleotide or polypeptide sequence to identify homology.

All publications and references, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference in their entirety as if each individual publication or reference were specifically and individually indicated to be incorporated by reference herein as being fully set forth. Any patent application to which this application claims priority is also incorporated by reference herein in its entirety in the manner described above for publications and references.

10 DEFINITIONS

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as the case may be, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heine, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., *SIAM J. Applied Math.*, 48: 1073 (1988). Methods to determine identity are designed to give the largest match between the sequences tested. Moreover, methods to determine identity are codified in publicly available computer programs. Computer program methods to determine identity between two sequences include, but are not limited to, the GAP program in the GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1):

387 (1984)), BLASTP, BLASTN (Altschul, S.F. et al., *J. Molec. Biol.* 215: 403-410 (1990), and FASTA(Pearson and Lipman Proc. Natl. Acad. Sci. USA 85; 2444-2448 (1988). The BLAST family of programs is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894; 5 Altschul, S., et al., *J. Mol. Biol.* 215: 403-410 (1990). The well known Smith Waterman algorithm may also be used to determine identity.

Parameters for polypeptide sequence comparison include the following:

Algorithm: Needleman and Wunsch, *J. Mol Biol.* 48: 443-453 (1970)

10 Comparison matrix: BLOSSUM62 from Henikoff and Henikoff, Proc. Natl. Acad. Sci. USA. 89:10915-10919 (1992)

Gap Penalty: 8

Gap Length Penalty: 2

A program useful with these parameters is publicly available as the "gap" program from 15 Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for peptide comparisons (along with no penalty for end gaps).

Parameters for polynucleotide comparison include the following:

Algorithm: Needleman and Wunsch, *J. Mol Biol.* 48: 443-453 (1970)

20 Comparison matrix: matches = +10, mismatch = 0

Gap Penalty: 50

Gap Length Penalty: 3

Available as: The "gap" program from Genetics Computer Group, Madison WI. These are the default parameters for nucleic acid comparisons.

25

A preferred meaning for "identity" for polynucleotides and polypeptides, as the case may be, are provided in (1) and (2) below.

(1) Polynucleotide embodiments further include an isolated polynucleotide comprising a polynucleotide sequence having at least a 50, 60, 70, 80, 85, 90, 95, 97 or 100% identity to the reference sequence of SEQ ID NO:1 or 3, wherein said polynucleotide sequence may be identical to the reference sequence of SEQ ID NO:1 or 3 or may include up to a certain integer number of nucleotide alterations as compared to the reference sequence, wherein said alterations are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence, and wherein said number of nucleotide alterations is determined by multiplying the total number of nucleotides in SEQ ID NO:1 or 3 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of nucleotides in SEQ ID NO:1 or 3, or:

$$n_n \leq x_n - (x_n \bullet y),$$

wherein n_n is the number of nucleotide alterations, x_n is the total number of nucleotides in SEQ ID NO:1 or 3, y is 0.50 for 50%, 0.60 for 60%, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%, 0.97 for 97% or 1.00 for 100%, and \bullet is the symbol for the multiplication operator, and wherein any non-integer product of x_n and y is rounded down to the nearest integer prior to subtracting it from x_n . Alterations of a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 or 4 may create nonsense, missense or frameshift mutations in this coding sequence and thereby alter the polypeptide encoded by the polynucleotide following such alterations.

By way of example, a polynucleotide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:1 or 3, that is it may be 100% identical, or it

may include up to a certain integer number of nucleic acid alterations as compared to the reference sequence such that the percent identity is less than 100% identity. Such alterations are selected from the group consisting of at least one nucleic acid deletion, substitution, including transition and transversion, or insertion, and wherein said

5 alterations may occur at the 5' or 3' terminal positions of the reference polynucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleic acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of nucleic acid alterations for a given percent identity is determined by multiplying the total number of nucleic acids in SEQ ID NO:1

10 or 3 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of nucleic acids in SEQ ID NO:1 or 3, or:

$$n_n \leq x_n - (x_n \bullet y),$$

15 wherein n_n is the number of nucleic acid alterations, x_n is the total number of nucleic acids in SEQ ID NO:1 or 3, y is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., \bullet is the symbol for the multiplication operator, and wherein any non-integer product of x_n and y is rounded down to the nearest integer prior to subtracting it from x_n .

20 (2) Polypeptide embodiments further include an isolated polypeptide comprising a polypeptide having at least a 50,60, 70, 80, 85, 90, 95, 97 or 100% identity to a polypeptide reference sequence of SEQ ID NO:2 or 4, wherein said polypeptide sequence may be identical to the reference sequence of SEQ ID NO:2 or 4 or may include up to a certain integer number of amino acid alterations as compared to the reference sequence,

25 wherein said alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal

positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence, and wherein said number of amino acid alterations is determined by multiplying the total number of amino acids in SEQ ID NO:2 or 4 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of amino acids in SEQ ID NO:2 or 4, or:

$$n_a \leq x_a - (x_a \bullet y),$$

wherein n_a is the number of amino acid alterations, x_a is the total number of amino acids in SEQ ID NO:2 or 4, y is 0.50 for 50%, 0.60 for 60%, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%, 0.97 for 97% or 1.00 for 100%, and \bullet is the symbol for the multiplication operator, and wherein any non-integer product of x_a and y is rounded down to the nearest integer prior to subtracting it from x_a .

By way of example, a polypeptide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:2 or 4, that is it may be 100% identical, or it may include up to a certain integer number of amino acid alterations as compared to the reference sequence such that the percent identity is less than 100% identity. Such alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of amino acid alterations for a given % identity is determined by multiplying the total number of amino acids in SEQ ID NO:2 or 4 by the integer defining the percent identity divided by 100 and

then subtracting that product from said total number of amino acids in SEQ ID NO:2 or 4, or:

$$n_a \leq x_a - (x_a \bullet y),$$

wherein n_a is the number of amino acid alterations, x_a is the total number of amino acids in SEQ ID NO:2 or 4, y is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., and \bullet is the symbol for the multiplication operator, and wherein any non-integer product of x_a and y is rounded down to the nearest integer prior to subtracting it from x_a .

"Individual(s)," when used herein with reference to an organism, means a multicellular eukaryote, including, but not limited to a metazoan, a mammal, an ovid, a bovid, a simian, a primate, and a human.

"Isolated" means altered "by the hand of man" from its natural state, *i.e.*, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein. Moreover, a polynucleotide or polypeptide that is introduced into an organism by transformation, genetic manipulation or by any other recombinant method is "isolated" even if it is still present in said organism, which organism may be living or non-living. Similarly, a polynucleotide or polypeptide whose expression is specifically altered by genetic manipulation is "isolated" even though the polynucleotide or polypeptide may be present in the organism in which it is naturally present.

"Polynucleotide(s)" generally refers to any polyribonucleotide or polydeoxyribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA including single and double-stranded regions.

- 5 "Variant" refers to a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes
- 10 may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical.
- 15 A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring
- 20 variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

- "Disease(s)" means any disease caused by or related to infection by bacteria, including, for example, otitis media, acute otitis media, recurrent otitis media, otitis media with
- 25 effusion, sinusitis, conjunctivitis, rhinopharyngitis, laryngitis, obstructive laryngitis, alveolitis, bronchitis, chronic bronchitis, enhancement of chronic obstructive pulmonary disease, complications of cystic fibrosis, pericarditis, endocarditis, osteomyelitis,

arthritis, genitourinary tract colonization and neonatal infection, bacteremia, septicemia, meningitis.

5

EXAMPLES:

The examples below are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. The examples are illustrative, but do not limit the invention.

10

Example 1: BASB070 gene from *Haemophilus influenzae* strain Rd KW20 and non-typeable *Haemophilus influenzae* (NTHi) strain 3224.

A: BASB070 in *Hi* strain Rd.

15

The BASB070 gene of SEQ ID NO:1 comes from *Haemophilus influenzae* strain Rd KW20. The translation of the BASB070 polynucleotide sequence is shown in SEQ ID NO:2.

B: BASB070 in *NTHi* strain 3224.

20

The sequence of the BASB070 gene comes from the sequencing of *NTHi* strain 3224.

Using the MegAlign program from the DNASTAR software package, an alignment of the polynucleotide sequences of SEQ ID NO:1 and 3 was performed, and is displayed in Figure 1; a pairwise comparison of identities shows that the two BASB070

25

polynucleotide gene sequences are 90.3 % identical. Using the same MegAlign program, an alignment of the polypeptide sequences of SEQ ID NO:2 and 4 was performed, and is displayed in Figure 2; a pairwise comparison of identities shows that the two BASB070 protein sequences are 90.6 % identical. These data show that the

BASB070 gene among the two strains NTHi 3224 and Hi RD are conserved but that there are also variable regions between them.

Example 2: Construction of Plasmid to Express Recombinant BASB070

5

A: Cloning of BASB070.

The *Nco*I and *Asp* 718 restriction sites CC ATG G and GG TAC C were engineered into specifically designed forward and reverse amplification primers, respectively, permitting directional cloning of a BASB070 PCR product into the commercially available *E. coli* expression plasmid pBADgIII(A) (Invitrogen, USA, ampicillin resistant). This plasmid provides the signal peptide from the bacteriophage fd pIII protein such that a mature BASB070 protein can be targeted to the periplasm of *E. coli*. The BASB070 PCR product was purified from the amplification reaction using Wizard PCR prepTM (Promega) according to the manufacturers instructions. To produce the required *Nco*I and *Asp* 718 termini necessary for cloning, purified PCR product was sequentially digested to completion with *Nco*I and *Asp* 718 restriction enzymes as recommended by the manufacturer (Boehringer Mannheim). Digested BASB070 PCR products and pBAD were gel-purified and ligated together using an approximately 5-fold molar excess of the digested fragment to the vector. A standard ~20 µl ligation reaction (~16°C, ~16 hours), using methods well known in the art, was performed using T4 DNA ligase (~2.0 units / reaction, Boehringer Mannheim). An aliquot of the ligation was used to transform electro-competent *E. coli* Top10 cells according to methods well known in the art. Following a ~2-3 hour outgrowth period at 37°C in ~1.0 ml of LB broth, transformed cells were plated on LB agar plates containing Ampicillin (50 µg/ml). Individual ampicillin-resistant colonies were selected and analyzed by whole cell-based PCR to verify that transformants contained the BASB070 DNA insert. Transformants that produced the expected PCR product were identified as

strains containing a BASB070 expression construct. Expression plasmid containing strains were then analyzed for the inducible expression of recombinant BASB070.

B: Expression Analysis of PCR-Positive Transformants.

- 5 For each PCR-positive transformant identified above, ~5.0 ml of LB broth containing ampicillin (50 µg/ml) was inoculated with cells from the patch plate and grown overnight at 37 °C with shaking (~250 rpm). An aliquot of the overnight seed culture (~1.0 ml) was inoculated into a 125 ml erlenmeyer flask containing ~25 ml of LB AMPICILLINE broth and grown at 37 °C with shaking (~250 rpm) until the culture
- 10 turbidity reached O.D.600 of ~0.5, i.e. mid-log phase (usually about 1.5 - 2.0 hours). At this time approximately half of the culture (~12.5 ml) was transferred to a second 125 ml flask and expression of recombinant BASB070 protein induced by the addition of L-Arabinose to a final concentration of 0.2 % (w/v). Incubation of both the arabinose-induced and non-induced cultures continued for an additional ~4 hours at 37 °C with
- 15 shaking. Samples (~1.0 ml) of both induced and non-induced cultures were removed after the induction period and the cells collected by centrifugation in a microcentrifuge at room temperature for ~3 minutes. Individual cell pellets were suspended in ~50µl of sterile water, then mixed with an equal volume of 2X Laemelli SDS-PAGE sample buffer containing 2-mercaptoethanol, and placed in boiling water bath for ~3 min to
- 20 denature protein. Equal volumes (~15µl) of both the crude arabinose-induced and the non-induced cell lysates were loaded onto duplicate 12% Tris/glycine polyacrylamide gel (1 mm thick Mini-gels, Novex). The induced and non-induced lysate samples were electrophoresed together with prestained molecular weight markers under conventional conditions using a standard SDS/Tris/glycine running buffer. Following
- 25 electrophoresis, one gel was stained with commassie brilliant blue R250 (BioRad) and then destained to visualize novel BASB070 arabinose-inducible protein(s).

NTHi Strains

The following strains of *Haemophilus influenzae* are provided as a useful reference for the present invention. The BASB070 gene utilised in accordance with the invention is not limited
5 with regard to the strain, but it may correspond to the BASB070 gene as found in any of the strains listed below or any related strain. This information is provided merely for convenience to those of skill in the art and is not an admission that any provision of a deposit is required for enablement.

strain 3219C (ET7)
10 strain 3241A (ET30)
strain 840645 (ET51)
strain 901905U (ET60)
strain A840177 (ET40)
strain A840177 (ET69)

15

All of the above strains were described in van Alphen, L., Caugant, D.A, Duim, B.A., O'Rourke, M., Bowler, L.D. (1997) Differences in genetic diversity of non-encapsulated *H. influenzae* from various diseases. Microbiology, 143: 1423-1431.

20 **HiRd Strains**

An example of a HiRd strain is described in R.D. Fleischmann et al., Science . Vol 269: 496-512 (1995) and K. W. Wilcox et al., J. Bact. Vol 122: 443 (1975) with the strain name KW20. This strain was deposited by the authors with the American Type Culture Collection under deposit
25 number ATCC 51907.

SEQUENCE INFORMATION

BASB070 Polynucleotide and Polypeptide Sequences

SEQ ID NO:1

5 *Haemophilus influenzae* BASB070 polynucleotide sequence from strain Rd KW20

ATGAAGAAAGCTATAAAATTAAATTTAATTACACTTGGCCTAATTAATACGATCGGTATGACGATTACACAAGCTCAAGC
 CGAAGAAACATTAGGACAAATTGATGTAGTGGAAGTATATCAAACGATAAAAAACCTTTCAGTGAAGCCAAAGCCA
 AAAGTACACGTGAAATGTCTTTAAGGAAACACAAACCATTGACCAAGTGATTGGAAGTATCCCTGGTGCATTTACTCAA
 CAAGATAAAGGCTCGGGTGTCTTTCTGTGAATATTCGTGGCGAAAATGGATTAGGTCGTGTCAATACTATGGTTGATGG
 10 TGTAAACACAGACCTTCTATTCTACAGCCTTAGACTCAGGTCAATCAGGCGGAAGTTCTCAATTTGGTGCGCAATCGATC
 CTAATTTTATTGCAGGTGTAGATGTTAATAAAGCAACTTTTCAGGAGCAAGCGGTATAAATGCGTTAGCAGGCAGTGCT
 AATTTTAGAACATTAGGCGTTAATGATGTTATTACCGATGACAAACCATTGGCATTATTCTGAAAGGAATGACAGGGAG
 TAATGCCACTAAATCCAATTTTATGACAATGGCTGCTGGCAGAAAATGGCTTGATAATGGTGGCTATGTAGGCGTGGTGT
 ATGGTTATAGCCAACGTGAAGTATCTCAAGATTACCGTATCGGTGGCGGAGAACGATTAGCATCATTAGGGCAGGATATT
 15 CTCGCGAAAGAAAAAGAAGCTTATTTTCGTAATGCGGGTTATATTTTAAATCCTGAAGGGCAATGGACACCTGATTTAAG
 CAAAAACATTGGTCTTGTAAACAACAGATTATCAGAAAAATGGTGATTGTAGTTATTATCGTATTGGATCTGCTGCAA
 AGACTAGAAGAGAAATTCTACAAGAATTATTAACAAATGGAAAAAACCTAAGGATATTGAAAAGCTCCAAAAAGGTAAT
 GATGGAATTGAAGAACTGACAAATCATTGAACGTAATAAAGATCAATATAGTGTGCACCGATTGAGCCGGGTAGTTT
 GCAATCTCGTTCTCGTAGCCATTTATTTAAATTTGAATATGGCGATGATCACCAAAATTTAGGGGCGCAATTACGCACGT
 20 TGGATAATAAAATTGGTTCTCGCAAAATTGAAAACCGTAATTACCAAGTCAATTATAAATTCAATAATAACAGCTATCTT
 GATCTTAATTTAATGGCTGCACATAACATTGGAAAACTATTTATCCTAAAGGCGGTTTTTTGTCTGGCTGGCAAGTGGC
 AGATAAACTTATCACTAAAAATGTCGCAATATTGTTGATATAACAACAGCCATACTTCTTACTGCCAAAAGAAATTG
 ATTTAAAAACCATAGGTTTAACTATTTTACCAATGAATACAGTAAAAACCGTTTTCCAGAAGAATTAAGTTTGT
 TATAACGATGCTTCACATGATCAAGGCTTATATTACACAGTAAAGAGGGCGATATTCTGGCAGAAAAAGTTTATTACC
 25 ACAACGTTTCAAGTAACTTACAACCTTCTGGCAAGCAAAATTTAAACCGTGTATTTTGATACCGCACTTTCTAAAGGCA
 TTTATCATTTAAATTACAGCGTGAATTTTACCCATTATGCCTTTAATGGTGAGTATGTAGGTTACGAAAATACAGCGGGT
 CAACAAATTAATGAACCTATTTTGCATAAATCAGGGCATAAAAAGGCATTCAATCATTCTGCCACATTAAGTGCAGAACT
 GAGTGATTATTTTATGCCATTTTCTTATTACGCACTCACAGAATGCCGAATATTCAAGAGATGTTTTCTCTCAAG
 TGTCTAATGCAGGGGTAACACAGCATTAAAACCTGAACAATCTGACACCTATCAACTAGGCTTTAATACTTATAAAAAA
 30 GGTCTCTTCACTCAAGACGATGTGCTAGGCGTAAATTAGTAGGCTATCGTAGCTTTATTAATACTATATCCATAATGT
 TTATGGTGTGGTGGCGAGATGGCATGCCTACGTGGGCAGAAAGTAATGGATTTAAATATACTATTGCTCATCAAAATT
 ATAAGCCTATTGTGAAAAGAGCGGCGTCGAGTTAGAAATTAATGACATGGGACGTTTTTTTGCGAATGTCTCTTAT
 GCATATCAACGAACAAATCAACCAACCAATTATGCCGATGCCAGCCGCGTCCGAATAATGCTTCACAAGAAGACATTTT
 GAAACAAGGTTATGGCTTATCTCGTGTTCATGCTACCAAAAGACTACGGCAGATTAGAGCTTGGCAGACGTTGGTTTG
 35 ATCAAAAATTAACCTTAGGCTGCGCAGCTCGTTATTATGGAAAAAGTAAACGTGCGACAATTGAAGAAGATATATCAAT
 GGATCTCGCTTTAAAAAAAATACCTTGCCTCGTGAAAATTAATGATGCGGTGAAAAAACGGAAGATATTAAAAACAACC
 GATTATTTTAGATTTACACGTCAGCTATGAACCAATCAAGATTTGATTATTAAGCGGAAGTACAAATCTATTAGATA
 AACGTTATGTTGATCGTTAGATGCTGGAATGACGCGGCTTCGCAACGTTATTATTCAAGTTTAAATAATCTATAGAA
 TGTGCGCAAGATTCTTCTGCTTGGGTGGTTCAGATAAAACCGTGCTTTATAACTTTGCACGTGGAAGAAGCTTATATTCT
 40 GAGTTTAACTATAAATCTAA

SEQ ID NO:2***Haemophilus influenzae* BASB070 polypeptide sequence deduced from the polynucleotide of SeQ ID NO:1**

5 MKKAIKLNLITLGLINTIGMTITQAQAEETLGQIDVVEKVISNDKKPFTEAKAKSTRENVFKETQTIDQVIRSIPGAFTQ
QDKGSGVSVNIRGENGLGRVNTMVDGVTQTFYSTALDSGQSGSSQFGAAIDPNFIAGVDVNKSNFSGASGINALAGSA
NFRTLGVNDVITDDKPFGLILKGMTGSNATKSNFMTMAAGRKWLNDNGYVGVVYGSQREVSQDYRIGGGERLASLGQDI
10 LAKEKEAYFRNAGYILNPEGQWTPDLSKKHWSCKNPDYQKNGDCSYRIGSAAKTRREILQELLTNGKKPKDIEKLQKGN
DGIETDKSFERNKDQYSVAPIEPGSLQSRSRSHLLKFEYGGDDHQNLAGQLRTLNDKIGSRKIENRNYQVNYNFNNNSYL
DLNLMAAHNIGKTIYPKGGFFAGWQVADKLITKNVANIVDINNSTFLLPKEIDLKTTLGFNYFTNEYSKNRFPPEELSLF
YNDASHDQGLYSHSKRGYSKSLLPQRSVILQPSGKQKFKTVYFDTALSKGIYHLNYSVNFTHYAFNGEYVGENTAG
QQINEPILHKS GHKAFNHSATLSAELSDYFMPFFTSRTHRMPNIQEMFFSQVSNAGVNTALKPEQSDTYQLGFNTYKK
GLFTQDDVLGVKLVGYRSFIKNYIHNVYGVWRDGMPTWAESNGFKYIAHQNYKPIVKKSGVELEINYDMGRFFANVSY
15 AYQRTNQPTNYADASPRNNASQEDILKQGYGLSRVSMPLPKDYGRLELGTWRFDQKLTGLAARYYGSKRATIEEEYIN
GSRFKNLTLRRENYAVKKTEDIKKQPIILDHVSYPEIKDLIIKAEVQNLLDKRYVDPLDAGNDAASQRYSSLNNSIE
CAQDSSACGGSDKTVLYNFARGRTYILSLNYKF

SEQ ID NO:3***Haemophilus influenzae* BASB070 polynucleotide sequence from strain ntHi 3224**

20 ATGAAGAAAGCTATAAAATTAAATTTAATTACACTTAGCCTAATCAATACAATCGGTATGACGATTACACAAGCTCAAGC
CGAAGAAACATTAGGGCAAATTGATGTCGTAGAAAAAGTGATATCAAAATGACAAAAACCTTTCACTGAAGCCAAAGCCA
AAAGTACGCGTGAAAATGTCTTTAAGGAAACACAAACCATTGACCAAGTCATTGCGAGCATTCTGCGGCATTTACTCAA
CAAGATAAAGGCTCGGGTGTGGTTTCTGTAAATATTCGTGGCGAAAATGGATTAGGTCTGTCAATACGATGGTTGATGG
25 TGTAACCCAAACCTTCTATTCTACAGCCTTAGACTCTGGTCAATCAGGCGGAAGTTCTCAATTTGGTGCAGCAATCGACC
CTAATTTTATGTCAGGTGTAGATGTTAATAAAGCAACTTTTCGGGAGCAAGCGGTATAAATGCCTTAGCAGGCAGTGCT
AATTTTAGAACATTAAAGCGTTAATGATGTGATTACCGATGACAAACCATTGCGCATTATTCTGAAAGGAATGACAGGGAG
CAATGCCACTAAATCCAATTTTATGACGACAGCTGCAGGCAGAAAATGGCTTGATAATGGTGGCTATGTAGGCGTAGTGT
ATGGTTATAGCCAACGTGAAGTTTCACAAGATTATCGTATAGGTGGCGGAGAACGATTAGCATCATTAGGGCAAGATATT
30 CTGCTAAAGAAAAAGAAAGATTTTCGTAATGATGGTTATGTTTTAAATCTGCTGGACAATGGGCACCTGATTTAAA
CAAACCACATTGGTCTTGTAATACCCGAGTTCTTTAAAGATAAAAGTATGAGTACATCTTGTAAGCCTTATCGTCTTG
GACCTGCTGCAACGACTAGACAAGAAATCTAAAAGAATTATTAGAAGATGGAAGAAACCTAAGGATATTGAAAAGCTC
CAAAAAGTAATGATGGAATTGAAGAACTGAAAAATCATTTGAACGTAATAAAGATCAATATGACGTCGCCCTATTGA
GCCTGGTAGTTTGCAATCTCGTTCACGTAGTCATTTATTAATTTGAATATAGCGATGATCACCATACGCTAGGGGCGC
35 AAATACGTACCCTTGATAATAAAATTGGTTCTCGCAAAATTGAAAACCGTAATTACCAAGTCAATTATAACTTCAATAAT
AACAGCTATCTTGATCTTAATTTAATGGCTGCACATAACATTGGCAAACTATTTATCCTAAGGGTGGTTTTTTTGCTGG
CTGGCAAGTGGCAGACAACTTATCACAAAAATGTGGCAATATTGTTGATATAAATAACAGCCATACTTTCTTACTGC
CAAAAGAAATCGATTTAAAACACATTAGGGTTTAACTATTTTACCAATGAATACAGTAAAAACCGTTTCCAGAAGAA
TTAAGTTTGTGTTTATGTGAATGAATCACATGATCAAGGCTTATATCACTCAGTAATAAAGGGCGATATCTGGCTCAAA
40 AGGTTTATTACCACAACGTTCAAGTAATCTTACAACCTTCTGGCAAGCAAAAAATTTAAACAGTGTATTTTGATACCGCAC
TTTCTAAAGGTATTTATCATTTAAATTACAGCGTGAATTTTACCCATTATGCCTTTAATGGTGAGTATGTTGGATATAAA
AATACAGCAGATAAAATTAATGAACCTATTTTGCATAAATCAGGGCATAAAAGGCATTCAATCATTCTGCTACATTAAG
TGCAGAGCTAAGTGATTATTTTATGCCATTTTTTACTTATTACGCACACACAGAATGCCGAATATTCAAGAGATGTTTT

TCTCTCAAGTGTCTGATGCTGGGGTAAACACCGCATTAAACCTGAACAATCTGACACCTATCAACTAGGCTTTAATACT
 TATAAAAAGGTCTATTCACTCAAGACGATGTATTAGGCATCAAATTAGTGGGCTATCGTAGCTTTATTA AAAACTATAT
 CCACAATGTGTATGGAGATTGGTCACGAGATGGTGTATGCCAGAGTGGGCAAGACTCAATGGTTTTCGTCTGACGATTG
 CTCATCAAAATTATCAACCAATAGTGA AAAAAGCGGAGCTGAGTTAGAGCTCAATTATGATATGGGGCGTTTTTTTGCA
 5 AATCTGTCTTATGCTTATCAACGTACTAATCAGCCAACCAATTATGCCGATGCCAGCTCACGTCCGCGTAATGCTTCAAA
 AGAAGAGATTTTGAAACAAGGTTATGGTTTATCAGCAATCTCTATGTTACCAAAGGACTACGGTAGATTAGAGCTTGGCA
 CACGCTGGTTTGATCAAAAATTAACCTCTTGGTATCGCAGCCCGTTACTATGGAAAAAGTAAACGTGCTACAACTCAAGAA
 GAATACATCAACGGCTCTCGCTATGAAAAAATACTACGCGCGACAGAATTTATTATGCTATTAAAAAGACAGAAGAGAT
 TAAAAACAACCTATTATTTTAGATTTACACGTCAGCTATGAACCAATCAAAGATTGATTATTAAAGCGGAAGTACAAA
 10 ATCTATTAGATAAACGTTATGTTGATCCGTTAGATGCTGGAAATGATGCGGCTTCGCAACGTTATTATTCAAGTTTAAAT
 GATTCTTTAGCCTGTAAATAAATGAATCAACCTGTAATGATGGTTCAGAGAAAAGTGTGCTTTATAACTTTGCACGTGG
 AAGAAGTATATTCTGAGTTTGAAGTATAAATCTAG

SEQ ID NO:4

15 *Haemophilus influenzae* BASB070 polypeptide sequence deduced from the
 polynucleotide of SeQ ID NO:3

MKKAIKLNLITLSLINTIGMTITQAQAEETLGQIDVVEKVISNDKKPFTEAKAKSTRENVFKETQTIDQVIRSIPGAFTQ
 QDKGSGVSVNIRGENGLGRVNTMVDGVTQTFYSTALDSGQSGSSQFGAAIDPNFIAGVDVNKSNFSGASGINALAGSA
 NFRTLSDNDVITDDKPFGIILKGMTGSNATKSNFMTTAAGRKWLNDNGGYGVVYGYSQREVSQDYRIGGGERLASLGQDI
 20 LAKEKEKIFRNDGYVLNSAGQWAPDLNKPWSCNTPSSLKDKSMSTSCPKYRLGPAATTRQEILKELLEDGKEPKDIEKL
 QKSNDGIEETEKSFERNKDQYDVAPIEPGSLQSRSRSHLLKFEYSDDHHTLGAQIRTLDNKIGSRKIENRNYQVNYNFNN
 NSYLDLNLMAAHNIGKTIYPKGGFFAGWQVADKLITKNVANIVDINNSHTFLLPKEIDLKTTLGFNYFTNEYSKNRFPPEE
 LSLFYVNESHDOGLYSLSNKGRYSGSKGLLPQRSVILQPSGKQKFKTVYFDTALSKGIYHLNYSVNFTHYAFNGEYVGYK
 NTADKINEPILHKSGHKKAFNHSATLSAELSDYFMPFFTYSRTHRMPNIQEMFFSQVSDAGVNTALKPEQSDTYQLGFNT
 25 YKKGLFTQDDVLGIKLVGYRSFIKNYIHNVDGWSRDGVMPPEARLNGFRLTIAHQNYQPIVKKSGAELELNYDMGRFFA
 NLSYAYQRTNQPTNYADASSRPRNASKEEILKQGYGLSRISMLPKDYGRLELGTWFDQKLTGLIAARYYGKSKRATTQE
 EYINGSRYEKNTTRDRIYYAIAKKTEEIKQPIILDHVSYPEIKDLIIKAEVQNLLDKRYVDPLDAGNDAASQRYSSLN
 DSLACKINESTCNDGSEKTVLYNFARGRTYILSLNYKF

CLAIMS

1. A vaccine composition comprising an effective amount of a polypeptide which polypeptide comprises an amino acid sequence which has at least 85% identity to the amino acid sequence of SEQ ID NO: 2 or 4 or to an immunogenic fragment thereof, or which polypeptide comprises a mimotope of the said amino acid sequence or immunogenic fragment, together with a pharmaceutically acceptable carrier.
2. A vaccine composition according to claim 1 wherein the amino acid sequence has at least 95% identity to the amino acid sequence of SEQ ID NO: 2 or 4 or to an immunogenic fragment thereof.
3. A vaccine composition comprising an effective amount of a polynucleotide which polynucleotide comprises a nucleotide sequence which has at least 85% identity to the nucleotide sequence of SEQ ID NO: 1, 3 or to a fragment thereof which encodes an immunogenic polypeptide, together with a pharmaceutically acceptable carrier.
4. The vaccine composition according to any one of claims 1 to 3 wherein said composition comprises at least one other *Haemophilus influenzae* antigen.
5. An expression vector or a recombinant live microorganism comprising an isolated polynucleotide which polynucleotide comprises a nucleotide sequence which has at least 85% identity to the amino acid sequence of SEQ ID NO: 1 or 3 or to a fragment thereof that encodes an immunogenic polypeptide.
6. A host cell comprising the expression vector of claim 5 or a membrane of said host cell expressing an isolated polypeptide comprising an amino acid sequence which

has at least 85% identity to the amino acid sequence of SEQ ID NO: 2 or 4, or to an immunogenic fragment thereof.

7. A process for producing a polypeptide comprising an amino acid sequence that has at least 85% identity to the amino acid sequence of SEQ ID NO: 2 or 4 or to an immunogenic fragment, comprising culturing a host cell of claim 6 under conditions sufficient for the production of said polypeptide and recovering the polypeptide from the culture medium.
8. A process for expressing a polynucleotide, which polynucleotide comprises a nucleotide sequence which has at least 85% identity to the nucleotide sequence of SEQ ID NO: 1 or 3 or to a fragment thereof that encodes an immunogenic polypeptide, the process comprising transforming a host cell with an expression vector comprising said polynucleotide and culturing said host cell under conditions sufficient for expression of said polynucleotide.
9. An antibody specific for the polypeptide of SEQ ID NO: 2 or 4 or an immunologically active fragment of the antibody.
10. A method of diagnosing a *Haemophilus influenzae* infection, comprising identifying a polypeptide which comprises an amino acid sequence which has at least 85% identity to the amino acid sequence of SEQ ID NO: 2 or 4 or a fragment thereof, or an antibody that is specific for said polypeptide, present within a biological sample from an animal suspected of having such an infection.
11. Use of a composition comprising an immunologically effective amount of a polypeptide which comprises an amino acid sequence which has at least 85% identity to the amino acid sequence of SEQ ID NO: 2 or 4 or to an immunogenic fragment thereof, or

which polypeptide comprises a mimotope of the said amino acid sequence or immunogenic fragment, in the preparation of a medicament for use in generating an immune response in a mammal.

- 5 12. Use of a composition comprising an immunologically effective amount of a polynucleotide which comprises a nucleotide sequence which has at least 85% identity to the nucleotide sequence of SEQ ID NO: 1 or 3 or to a fragment thereof that encodes an immunogenic polypeptide, in the preparation of a medicament for use in generating an immune response in a mammal.

10

13. A therapeutic composition useful in treating humans with *Haemophilus influenzae* disease comprising at least one antibody directed against the polypeptide of SEQ ID NO: 2 or 4 and a suitable pharmaceutical carrier.

- 15 14. An isolated polypeptide comprising the amino acid sequence of SEQ ID NO: 4 or a fragment or a mimotope thereof.

15. An isolated polynucleotide comprising a nucleotide sequence encoding the polypeptide of claim 14.

20

16. The polynucleotide of claim 15 comprising the nucleotide sequence of SEQ ID NO: 3 a fragment thereof.

Figure 1 : Alignment of the BASB070 polynucleotide sequences.
Identity to SeqID No:1 is indicated by a dot and Gap is indicated by a dash.

5		*	20	*	
	Seqid1	:	ATGAAGAAAGCTATAAAATTAAATTTAATT	:	30
	Seqid3	:	:	30
10			40	*	60
	Seqid1	:	ACACTTGGCCTAATTAATACGATCGGTATG	:	60
	Seqid3	:A.....C.....A.....	:	60
15		*	80	*	
	Seqid1	:	ACGATTACACAAGCTCAAGCCGAAGAAACA	:	90
	Seqid3	:	:	90
20			100	*	120
	Seqid1	:	TTAGGACAAATTGATGTAGTGGAAAAAGTT	:	120
	Seqid3	:G.....C..A.....G	:	120
25		*	140	*	
	Seqid1	:	ATATCAAACGATAAAAAACCTTTCCTGAA	:	150
	Seqid3	:T..C.....	:	150
30			160	*	180
	Seqid1	:	GCCAAAGCCAAAAGTACACGTGAAAATGTC	:	180
	Seqid3	:G.....	:	180

2 / 21

```

                *           200           *
Seqid1 : TTTAAGGAAACACAAACCATTGACCAAGTG : 210
Seqid3 : .....C : 210

5
                220           *           240
Seqid1 : ATTCGAAGTATCCCTGGTGCATTTACTCAA : 240
Seqid3 : .....G..C..T.....G..... : 240

10
                *           260           *
Seqid1 : CAAGATAAAGGCTCGGGTGTCTGTTTCTGTG : 270
Seqid3 : .....G.....A : 270

15
                280           *           300
Seqid1 : AATATTCGTGGCGAAAATGGATTAGGTCGT : 300
Seqid3 : ..... : 300

20
                *           320           *
Seqid1 : GTCAATACTATGGTTGATGGTGTAAACACAG : 330
Seqid3 : .....G.....C..A : 330

25
                340           *           360
Seqid1 : ACCTTCTATTCTACAGCCTTAGACTCAGGT : 360
Seqid3 : .....T... : 360

30
                *           380           *
Seqid1 : CAATCAGGCGGAAGTTCTCAATTTGGTGCG : 390
Seqid3 : ..... : 390
```

5

Seqid1 : GCAATCGATCCTAATTTTATTGCAGGTGTA : 420
Seqid3 :C..... : 420

10

Seqid1 : GATGTTAATAAAAGCAACTTTTCAGGAGCA : 450
Seqid3 :G..... : 450

15

Seqid1 : AGCGGTATAAATGCGTTAGCAGGCAGTGCT : 480
Seqid3 :C..... : 480

20

Seqid1 : AATTTTAGAACATTAGGCGTTAATGATGTT : 510
Seqid3 :A.....G : 510

25

Seqid1 : ATTACCGATGACAAACCATTGTCATTATT : 540
Seqid3 :C..... : 540

30

Seqid1 : CTGAAAGGAATGACAGGGAGTAATGCCACT : 570
Seqid3 :C..... : 570

Seqid1 : AAATCCAATTTTATGACAATGGCTGCTGGC : 600
Seqid3 :G.CA....A... : 600

```

                                *           620           *
Seqid1 : AGAAAATGGCTTGATAATGGTGGCTATGTA : 630
5 Seqid3 : ..... : 630

                                640           *           660
Seqid1 : GGCGTGGTGTATGGTTATAGCCAACGTGAA : 660
10 Seqid3 : .....A..... : 660

                                *           680           *
Seqid1 : GTATCTCAAGATTACCGTATCGGTGGCGGA : 690
15 Seqid3 : ..T..A.....T.....A..... : 690

                                700           *           720
Seqid1 : GAACGATTAGCATCATTAGGGCAGGATATT : 720
20 Seqid3 : .....A..... : 720

                                *           740           *
Seqid1 : CTCGCGAAAGAAAAAGAAGCTTATTTTCGT : 750
25 Seqid3 : ..T..T.....AAGAT..... : 750

                                760           *           780
Seqid1 : AATGCGGGTTATATTTTAAATCCTGAAGGG : 780
30 Seqid3 : ....AT.....G.....T...CT..A : 780

                                *           800           *
Seqid1 : CAATGGACACCTGATTTAAGCAAAAACAT : 810
```

```

Seqid3 : .....G.....A....CC.... : 810

                        820          *          840
5  Seqid1 : TGGTCTTGTAACAAACCAGATTATCAGAAA : 840
   Seqid3 : .....T.CC..GAG..C.TTA... : 840

                        *          860          *
10 Seqid1 : AATGGTGAT-----TGTTAGTTAT : 858
   Seqid3 : G..AAAAG.ATGAGTACATCT....AGCC. : 870

                        880          *          900
15 Seqid1 : TATCGTATTGGATCTGCTGCAAAGACTAGA : 888
   Seqid3 : .....C.....C.....C..... : 900

20

                        *          920          *
   Seqid1 : AGAGAAATTCTACAAGAATTATTAACAAAT : 918
   Seqid3 : CA.....A.....GA.G.. : 930

25

                        940          *          960
   Seqid1 : GGAAAAAACCTAAGGATATTGAAAAGCTC : 948
   Seqid3 : .....G..... : 960

30

                        *          980          *
   Seqid1 : CAAAAAGGTAATGATGGAATTGAAGAACT : 978
   Seqid3 : .....A..... : 990

```

1000 * 1020
Seqid1 : GACAAATCATTGTAACGTAATAAAGATCAA : 1008
Seqid3 : ..A..... : 1020
5

* 1040 *
Seqid1 : TATAGTGTGACCCGATTGAGCCGGGTAGT : 1038
Seqid3 : ...GAC..C..C..T.....T..... : 1050
10

1060 * 1080
Seqid1 : TTGCAATCTCGTTCTCGTAGCCATTTATTA : 1068
Seqid3 :A.....T..... : 1080
15

* 1100 *
Seqid1 : AAATTTGAATATGGCGATGATCACCAAAT : 1098
Seqid3 :A.....T.CG : 1110
20

1120 * 1140
Seqid1 : TTAGGGGCGCAATTACGCACGTTGGATAAT : 1128
Seqid3 : C.....A....T..CC.T..... : 1140
25

* 1160 *
Seqid1 : AAAATTGGTTCTCGCAAATTGAAAACCGT : 1158
Seqid3 : : 1170
30

1180 * 1200
Seqid1 : AATTACCAAGTCAATTATAACTTCAATAAT : 1188
Seqid3 : : 1200

```

                                *           1220           *
Seqid1 : AACAGCTATCTTGATCTTAATTTAATGGCT : 1218
5 Seqid3 : ..... : 1230

                                1240           *           1260
Seqid1 : GCACATAACATTGGAAAACTATTTATCCT : 1248
10 Seqid3 : .....C..... : 1260

                                *           1280           *
Seqid1 : AAAGGCGGTTTTTTTGCTGGCTGGCAAGTG : 1278
15 Seqid3 : ..G..T..... : 1290

                                1300           *           1320
Seqid1 : GCAGATAAACTTATCACTAAAAATGTGCGCA : 1308
20 Seqid3 : .....C.....A.....G... : 1320

                                *           1340           *
Seqid1 : AATATTGTTGATATAAACACAGCCATACT : 1338
25 Seqid3 : .....T..... : 1350

                                1360           *           1380
Seqid1 : TTCTTACTGCCAAAAGAAATTGATTTAAAA : 1368
30 Seqid3 : .....C..... : 1380

                                *           1400           *
Seqid1 : ACCACATTAGGTTTTAACTATTTTACCAAT : 1398
```

Seqid3 :G..... : 1410

1420 * 1440

5 Seqid1 : GAATACAGTAAAAACCGTTTTCAGAAAGAA : 1428

Seqid3 : : 1440

* 1460 *

10 Seqid1 : TTAAGTTTGTTTTATAACGATGCTTCACAT : 1458

Seqid3 :GTGA...AA..... : 1470

1480 * 1500

15 Seqid1 : GATCAAGGCTTATATTACACAGTAAAAGA : 1488

Seqid3 :T.....T.A. : 1500

* 1520 *

20 Seqid1 : GGGCGATATTCTGGCACAAAAAGTTTATTA : 1518

Seqid3 :T.....G..... : 1530

1540 * 1560

25 Seqid1 : CCACAACGTTTCAGTAATCTTACAACCTTCT : 1548

Seqid3 : : 1560

* 1580 *

30 Seqid1 : GGCAAGCAAAAATTTAAAACCGTGTATTTT : 1578

Seqid3 :A..... : 1590

5
Seqid1 : 1600 * 1620
GATACCGCACTTTCTAAAGGCATTTATCAT : 1608
Seqid3 :T..... : 1620

10
Seqid1 : * 1640 *
TTAAATTACAGCGTGAATTTTACCCATTAT : 1638
Seqid3 : : 1650

15
Seqid1 : 1660 * 1680
GCCTTTAATGGTGAGTATGTAGGTTACGAA : 1668
Seqid3 :T..A..TA.. : 1680

20
Seqid1 : * 1700 *
AATACAGCGGGTCAACAAATTAATGAACCT : 1698
Seqid3 :A.A.---A..... : 1707

25
Seqid1 : 1720 * 1740
ATTTTGCATAAATCAGGGCATAAAAAGGCA : 1728
Seqid3 : : 1737

30
Seqid1 : * 1760 *
TTCAATCATTCTGCCACATTAAGTGCAGAA : 1758
Seqid3 :T.....G : 1767

Seqid1 : 1780 * 1800
CTGAGTGATTATTTTATGCCATTTTTTACT : 1788
Seqid3 : ..A..... : 1797

```

                                *           1820           *
Seqid1 : TATTCACGCACTCACAGAATGCCGAATATT : 1818
5 Seqid3 : .....A..... : 1827

                                1840           *           1860
Seqid1 : CAAGAGATGTTTTTCTCTCAAGTGTCTAAT : 1848
10 Seqid3 : .....G.. : 1857

                                *           1880           *
Seqid1 : GCAGGGGTAAACACAGCATTTAAACCTGAA : 1878
15 Seqid3 : ..T.....C..... : 1887

                                1900           *           1920
Seqid1 : CAATCTGACACCTATCAACTAGGCTTTAAT : 1908
20 Seqid3 : ..... : 1917

                                *           1940           *
Seqid1 : ACTTATAAAAAAGGTCTCTTCACTCAAGAC : 1938
25 Seqid3 : .....A..... : 1947

                                1960           *           1980
Seqid1 : GATGTGCTAGGCGTAAAATTAGTAGGCTAT : 1968
30 Seqid3 : .....AT.....A.C.....G..... : 1977

                                *           2000           *
Seqid1 : CGTAGCTTTATTAAAACTATATCCATAAT : 1998
```

Seqid3 :C... : 2007

2020 * 2040

5 Seqid1 : GTTTATGGTGTGTTGGTGGCGAGATGGC--- : 2025
Seqid3 : ..G.....A.A.....CA.....TGTT : 2037

* 2060 *

10 Seqid1 : ATGCCTACGTGGGCAGAAAGTAATGGATTT : 2055
Seqid3 :AGA.....AG.CTC.....T... : 2067

2080 * 2100

15 Seqid1 : AAATATACTATTGCTCATCAAATTATAAG : 2085
Seqid3 : CGTCTG..G.....C.A : 2097

* 2120 *

20 Seqid1 : CCTATTGTGAAAAAGAGCGGCGTCGAGTTA : 2115
Seqid3 : ..A..A.....A.....A.CT..... : 2127

2140 * 2160

25 Seqid1 : GAAATTA ACTATGACATGGGACGTTTTTTT : 2145
Seqid3 : ..GC.C..T.....T.....G..... : 2157

* 2180 *

30 Seqid1 : GCGAATGTCTCTTATGCATATCAACGAACA : 2175
Seqid3 : ..A...C.G.....T.....T..T : 2187

2200 * 2220

```

*                2240                *
Seqid1  :  CCGCGTCCGAATAATGCTTCACAAGAAGAC  :  2235
Seqid3  :  T.A.....CG.....A.....G      :  2247

```

```

                2260                *                2280
Seqid1  :  ATTTTGAAACAAGGTTATGGCTTATCTCGT  :  2265
Seqid3  :  .....T.....A..A  :  2277

```

```

*                2300                *
Seqid1  :  GTTTCAATGCTACCAAAGACTACGGCAGA  :  2295
Seqid3  :  A.C..T...T.....G.....T...  :  2307

```

```

                2320                *                2340
Seqid1  :  TTAGAGCTTGGCACACGTTGGTTTGATCAA  :  2325
Seqid3  :  .....C.....  :  2337

```

```

*                2360                *
Seqid1  :  AAATTAACCTTAGGTCTGGCAGCTCGTTAT  :  2355
Seqid3  :  .....TC.T...A.C.....C.....C  :  2367

```

```

                2380                *                2400
Seqid1  :  TATGGAAAAAGTAAACGTGCGACAATTGAA  :  2385
Seqid3  :  .....T....C.C..  :  2397

```

```

                *           2420           *
Seqid1 : GAAGAATATATCAATGGATCTCGCTTTAAA : 2415
5 Seqid3 : .....C.....C..C.....A.G.. : 2427

                2440           *           2460
Seqid1 : AAAAATACCTTGCGTCGTGAAAATTACTAT : 2445
10 Seqid3 : .....TAC...CGACAG..T...T... : 2457

                *           2480           *
Seqid1 : GCCGTGAAAAAAACGGAAGATATTAAAAAA : 2475
15 Seqid3 : ..TA.T.....G..A.....G..... : 2487

                2500           *           2520
Seqid1 : CAACCGATTATTTTAGATTTACACGTCAGC : 2505
20 Seqid3 : .....T..... : 2517

                *           2540           *
Seqid1 : TATGAACCAATCAAAGATTTGATTATTAAA : 2535
25 Seqid3 : ..... : 2547

                2560           *           2580
Seqid1 : GCGGAAGTACAAAATCTATTAGATAAACGT : 2565
30 Seqid3 : ..... : 2577

                *           2600           *
Seqid1 : TATGTTGATCCGTTAGATGCTGGAAATGAC : 2595
```

Seqid3 :T : 2607

2620 * 2640

5 Seqid1 : GCGGCTTCGCAACGTTATTATTCAAGTTTA : 2625

Seqid3 : : 2637

* 2660 *

10 Seqid1 : AATAATTCTATAGAATGTGCGCAAGATTCT : 2655

Seqid3 : ...G.....T...CC...AAAAT.A..GAA : 2667

2680 * 2700

15 Seqid1 : TCTGCTTGC---GGTGGTTCAGATAAAACC : 2682

Seqid3 : ..AA.C..TAAT.A.....G.....T : 2697

* 2720 *

20 Seqid1 : GTGCTTTATAACTTTGCACGTGGAAGAACT : 2712

Seqid3 : : 2727

2740 * 2760

25 Seqid1 : TATATTCTGAGTTTAAACTATAAATTCTAA : 2742

Seqid3 :G.....G : 2757

Figure 2 : Alignment of the BASB070 polypeptide sequences.
Identity to SeqID No:2 is indicated by a dot and Gap is indicated by a dash.

```

                *           20           *
5  Seqid2 : MKKAIKLNLITLGLINTIGMTITQAQAEET : 30
    Seqid4 : .....S..... : 30

                40           *           60
10 Seqid2 : LGQIDVVEKVISNDKKPFTEAKAKSTRENV : 60
    Seqid4 : ..... : 60

                *           80           *
15 Seqid2 : FKETQTIDQVIRSIPGAFTQQDKGSGVVSV : 90
    Seqid4 : ..... : 90

                100           *           120
20 Seqid2 : NIRGENGLGRVNTMVDGVTQTFYSTALDSG : 120
    Seqid4 : ..... : 120

                *           140           *

```

16 / 21

Seqid2 : QSGGSSQFGAAIDPNFIAGVDVNKSNFSGA : 150

Seqid4 : : 150

5

160

*

180

Seqid2 : SGINALAGSANFRTLGVNDVITDDKPFGII : 180

Seqid4 :S..... : 180

10

*

200

*

Seqid2 : LKGMTGSNATKSNFMTMAAGRKWLDNGGYV : 210

Seqid4 :T..... : 210

15

220

*

240

Seqid2 : GVVYGYSQREVSQDYRIGGGERLASLGQDI : 240

Seqid4 : : 240

20

*

260

*

Seqid2 : LAKEKEAYFRNAGYILNPEGQWTPDLSKKH : 270

Seqid4 :KI...D.....SA...A...N.P. : 270

17 / 21

280 * 300

Seqid2 : WSCNKPDYQKN----GDCSYRIGSAAKTR : 296

Seqid4 :T.SSL..KSMSTS.KP....P..T.. : 300

5

* 320 *

Seqid2 : REILQELLTNGKKPKDIEKLQKGNDGIEET : 326

Seqid4 : Q...K...E...E.....S..... : 330

10

340 * 360

Seqid2 : DKSFERNKDQYSVAPIEPGSLQSRSRSHLL : 356

Seqid4 : E.....D..... : 360

15

* 380 *

Seqid2 : KFEYGDDHQNLGAQLRTLDNKIGSRKIENR : 386

Seqid4 :S...HT..... : 390

20

400 * 420

Seqid2 : NYQVNYNFNNNSYLDLNLMAAHNIGKTIYP : 416

Seqid4 : : 420

* 440 *

5 Seqid2 : KGGFFAGWQVADKLITKNVANIVDINNSHT : 446

Seqid4 : : 450

460 * 480

10 Seqid2 : FLLPKEIDLKTTTLGFNYFTNEYSKNRFPEE : 476

Seqid4 : : 480

* 500 *

15 Seqid2 : LSLFYNDASHDQGLYSHSKRGRYSGTKSLL : 506

Seqid4 :V.E.....L.N.....G.. : 510

520 * 540

20 Seqid2 : PQRSVILQPSGKQKFCTVYFDTALSKGIYH : 536

Seqid4 : : 540

19 / 21

* 560 *

Seqid2 : LNYSVNFTHYAFNGEYVGYENTAGQQINEP : 566

Seqid4 :K...-DK.... : 569

5

580 * 600

Seqid2 : ILHKSGHKKAFNHSATLSAELSDYFMPFFT : 596

Seqid4 : : 599

10

* 620 *

Seqid2 : YSRTHRMPNIQEMFFSQVSNAGVNTALKPE : 626

Seqid4 : : 629

15

640 * 660

Seqid2 : QSDTYQLGFNTYKKGLFTQDDVLGVKLVGY : 656

Seqid4 : : 659

20

* 680 *

Seqid2 : RSFIKNYIHNVYGVWWRDG-MPTWAESNGF : 685

Seqid4 :D.S...V..E..RL... : 689

20 / 21

700 * 720

Seqid2 : KYTIAHQNYKPIVKKSGVELEINYDMGRFF : 715

Seqid4 : .L.....Q.....A..... : 719

5

* 740 *

Seqid2 : ANVSYAYQRTNQPTNYADASPRPNNASQED : 745

Seqid4 :S..R...K.E : 749

10

760 * 780

Seqid2 : ILKQGYGLSRVSMPLPKDYGRLELGTRWFDQ : 775

Seqid4 : : 779

15

* 800 *

Seqid2 : KLTLGLAARYYGKSKRATIEEEYINGSRFX : 805

Seqid4 :T.....E : 809

20

820 * 840

Seqid2 : KNTLRRENYAVKKTEDIKKQPIILDLHVS : 835

Seqid4 : ...T.DRI.....E..... : 839

* 860 *

5 Seqid2 : YEPIKDLIIKAEVQNLLDKRYVDPLDAGND : 865

Seqid4 : : 869

880 * 900

10 Seqid2 : AASQRYYYSSLNNSIECAQDSSACG-GSDKT : 894

Seqid4 :A.KI.E.T.ND..E.. : 899

*

15 Seqid2 : VLYNFARGRTYILSLNYKF : 913

Seqid4 : : 918

SEQUENCE LISTING

<110> SmithKline Beecham Biologicals S.A.

5

<120> Novel Compounds

<130> BM45376

10

<160> 4

<170> FastSEQ for Windows Version 3.0

15

<210> 1

<211> 2742

<212> DNA

<213> *Haemophilus influenzae* strain Rd KW20

20

<400> 1

	atgaagaaag ctataaaatt aaattttaatt acacttggcc taattaatac gatcgggtatg	60
	acgattacac aagctcaagc cgaagaaaca ttaggacaaa ttgatgtagt ggaaaaagtt	120
	atatcaaacg ataaaaaaccc ttctactgaa gccaaagcca aaagtacacg tgaaaatgtc	180
	tttaaggaaa cacaaacccat tgaccaagtgt attcgaagta tccctgggtgc atttactcaa	240
25	caagataaag gctcgggtgt cgtttctgtg aatattcgtg gcgaaaatgg attaggtcgt	300
	gtcaatacta tggttgatgg tgtaacacag accttctatt ctacagcctt agactcaggt	360
	caatcaggcg gaagttctca atttggtgcg gcaatcgatc ctaattttat tgcagggtga	420
	gatgttaata aaagcaactt ttcaggagca agcgggtataa atgcgttagc aggcagtgct	480
	aatttttagaa cattaggcgt taatgatgtt attaccgatg acaaaccatt tggcattatt	540
30	ctgaaaggaa tgacagggag taatgccact aaatccaatt ttatgacaat ggctgctggc	600
	agaaaatggc ttgataatgg tggctatgta ggcgtggtgt atggttatag ccaacgtgaa	660
	gtatctcaag attaccgtat cggtgggcga gaacgattag catcattagg gcaggatatt	720
	ctcgcgaaag aaaaagaagc ttattttcgt aatgcgggtt atattttaaa tcctgaaggg	780
	caatggacac ctgatttaag caaaaaacat tggctctgta acaaaccaga ttatcagaaa	840
35	aatggtgatt gtagttatta tcgtattgga tctgctgcaa agactagaag agaaattcta	900
	caagaattat taacaaatgg aaaaaaacct aaggatattg aaaagctcca aaaaggtaat	960
	gatggaattg aagaaactga caaatcattt gaacgtaata aagatcaata tagtggtgca	1020
	ccgattgagc cgggtagttt gcaatctcgt tctcgtagcc atttattaaa atttgaatat	1080
	ggcgatgatc accaaaattt aggggcgcaa ttacgcacgt tggataataa aattggttct	1140
40	cgcaaaattg aaaaccgtaa ttaccaagtc aattataact tcaataataa cagctatctt	1200

```

gatcttaatt taatggctgc acataacatt ggaaaaacta tttatcctaa aggcggtttt 1260
tttgctggct ggcaagtggc agataaaactt atcactaaaa atgtcgcaaa tattgttgat 1320
ataaacaaca gccatacttt cttactgcca aaagaaattg atttaaaaac cacattaggt 1380
tttaactatt ttaccaatga atacagtaaa aaccgttttc cagaagaatt aagtttgttt 1440
5 tataacgatg cttcacatga tcaaggctta tattcacaca gtaaaagagg gcgatattct 1500
ggcacaaaaa gtttattacc acaacgttca gtaatcttac aaccttctgg caagcaaaaa 1560
tttaaaaccg tgtattttga taccgcactt tctaaaggca tttatcattt aaattacagc 1620
gtgaatttta cccattatgc ctttaatggg gagtatgtag gttacgaaaa tacagcgggt 1680
caacaaatta atgaacctat ttgcataaaa tcagggcata aaaaggcatt caatcattct 1740
10 gccacattaa gtgcagaact gagtgattat tttatgccat tttttactta ttcacgcact 1800
cacagaatgc cgaatattca agagatgttt ttctctcaag tgtctaattg aggggtaaac 1860
acagcattaa aacctgaaca atctgacacc tatcaactag gctttaatac ttataaaaaa 1920
ggtctcttca ctcaagacga tgtgctaggc gtaaaattag taggctatcg tagctttatt 1980
aaaaactata tccataatgt ttatgggtgt tggtggcgag atggcatgcc tacgtgggca 2040
15 gaaagtaatg gatttaataa tactattgct catcaaaatt ataagcctat tgtgaaaaag 2100
agcggcgctg agttagaaat taactatgac atgggacgtt tttttgcgaa tgtctcttat 2160
gcatatcaac gaacaaatca accaaccaat tatgccgatg ccagcccgcg tccgaataat 2220
gcttcacaag aagacatttt gaaacaaggt tatggcttat ctcgtgtttc aatgctacca 2280
aaagactacg gcagattaga gcttggcaca cgttggtttg atcaaaaatt aaccttaggt 2340
20 ctggcagctc gttattatgg aaaaagtaaa cgtgcgacaa ttgaagaaga atatatcaat 2400
ggatctcgct ttaaaaaaaaa taccttgctg cgtgaaaatt actatgccgt gaaaaaacg 2460
gaagatatta aaaaacaacc gattatttta gatttacacg tcagctatga accaatcaaa 2520
gatttgatta ttaaagcgga agtacaaaat ctattagata aacgttatgt tgatccgtta 2580
gatgctggaa atgacgcggc ttgcgaacgt tattattcaa gtttaaataa ttctatagaa 2640
25 tgtgcgcaag attcttctgc ttgcgggtgt tcagataaaa ccgtgcttta taactttgca 2700
cgtggaagaa cttatattct gagtttaaac tataaattct aa 2742

```

<210> 2

<211> 913

30 <212> PRT

<213> *Haemophilus influenzae* strain Rd KW20

<400> 2

```

Met Lys Lys Ala Ile Lys Leu Asn Leu Ile Thr Leu Gly Leu Ile Asn
35 1 5 10 15
Thr Ile Gly Met Thr Ile Thr Gln Ala Gln Ala Glu Glu Thr Leu Gly
20 25 30
Gln Ile Asp Val Val Glu Lys Val Ile Ser Asn Asp Lys Lys Pro Phe
35 40 45
40 Thr Glu Ala Lys Ala Lys Ser Thr Arg Glu Asn Val Phe Lys Glu Thr
50 55 60
Gln Thr Ile Asp Gln Val Ile Arg Ser Ile Pro Gly Ala Phe Thr Gln
65 70 75 80

```

Gln Asp Lys Gly Ser Gly Val Val Ser Val Asn Ile Arg Gly Glu Asn
 85 90 95
 Gly Leu Gly Arg Val Asn Thr Met Val Asp Gly Val Thr Gln Thr Phe
 100 105 110
 5 Tyr Ser Thr Ala Leu Asp Ser Gly Gln Ser Gly Gly Ser Ser Gln Phe
 115 120 125
 Gly Ala Ala Ile Asp Pro Asn Phe Ile Ala Gly Val Asp Val Asn Lys
 130 135 140
 10 Ser Asn Phe Ser Gly Ala Ser Gly Ile Asn Ala Leu Ala Gly Ser Ala
 145 150 155 160
 Asn Phe Arg Thr Leu Gly Val Asn Asp Val Ile Thr Asp Asp Lys Pro
 165 170 175
 Phe Gly Ile Ile Leu Lys Gly Met Thr Gly Ser Asn Ala Thr Lys Ser
 180 185 190
 15 Asn Phe Met Thr Met Ala Ala Gly Arg Lys Trp Leu Asp Asn Gly Gly
 195 200 205
 Tyr Val Gly Val Val Tyr Gly Tyr Ser Gln Arg Glu Val Ser Gln Asp
 210 215 220
 Tyr Arg Ile Gly Gly Gly Glu Arg Leu Ala Ser Leu Gly Gln Asp Ile
 20 225 230 235 240
 Leu Ala Lys Glu Lys Glu Ala Tyr Phe Arg Asn Ala Gly Tyr Ile Leu
 245 250 255
 Asn Pro Glu Gly Gln Trp Thr Pro Asp Leu Ser Lys Lys His Trp Ser
 260 265 270
 25 Cys Asn Lys Pro Asp Tyr Gln Lys Asn Gly Asp Cys Ser Tyr Tyr Arg
 275 280 285
 Ile Gly Ser Ala Ala Lys Thr Arg Arg Glu Ile Leu Gln Glu Leu Leu
 290 295 300
 Thr Asn Gly Lys Lys Pro Lys Asp Ile Glu Lys Leu Gln Lys Gly Asn
 30 305 310 315 320
 Asp Gly Ile Glu Glu Thr Asp Lys Ser Phe Glu Arg Asn Lys Asp Gln
 325 330 335
 Tyr Ser Val Ala Pro Ile Glu Pro Gly Ser Leu Gln Ser Arg Ser Arg
 340 345 350
 35 Ser His Leu Leu Lys Phe Glu Tyr Gly Asp Asp His Gln Asn Leu Gly
 355 360 365
 Ala Gln Leu Arg Thr Leu Asp Asn Lys Ile Gly Ser Arg Lys Ile Glu
 370 375 380
 Asn Arg Asn Tyr Gln Val Asn Tyr Asn Phe Asn Asn Asn Ser Tyr Leu
 40 385 390 395 400
 Asp Leu Asn Leu Met Ala Ala His Asn Ile Gly Lys Thr Ile Tyr Pro
 405 410 415
 Lys Gly Gly Phe Phe Ala Gly Trp Gln Val Ala Asp Lys Leu Ile Thr

	420		425		430
	Lys Asn Val Ala Asn Ile Val Asp Ile Asn Asn Ser His Thr Phe Leu				
	435		440		445
5	Leu Pro Lys Glu Ile Asp Leu Lys Thr Thr Leu Gly Phe Asn Tyr Phe				
	450		455		460
	Thr Asn Glu Tyr Ser Lys Asn Arg Phe Pro Glu Glu Leu Ser Leu Phe				
	465		470		475
	Tyr Asn Asp Ala Ser His Asp Gln Gly Leu Tyr Ser His Ser Lys Arg				
	485		490		495
10	Gly Arg Tyr Ser Gly Thr Lys Ser Leu Leu Pro Gln Arg Ser Val Ile				
	500		505		510
	Leu Gln Pro Ser Gly Lys Gln Lys Phe Lys Thr Val Tyr Phe Asp Thr				
	515		520		525
	Ala Leu Ser Lys Gly Ile Tyr His Leu Asn Tyr Ser Val Asn Phe Thr				
15	530		535		540
	His Tyr Ala Phe Asn Gly Glu Tyr Val Gly Tyr Glu Asn Thr Ala Gly				
	545		550		555
	Gln Gln Ile Asn Glu Pro Ile Leu His Lys Ser Gly His Lys Lys Ala				
	565		570		575
20	Phe Asn His Ser Ala Thr Leu Ser Ala Glu Leu Ser Asp Tyr Phe Met				
	580		585		590
	Pro Phe Phe Thr Tyr Ser Arg Thr His Arg Met Pro Asn Ile Gln Glu				
	595		600		605
	Met Phe Phe Ser Gln Val Ser Asn Ala Gly Val Asn Thr Ala Leu Lys				
25	610		615		620
	Pro Glu Gln Ser Asp Thr Tyr Gln Leu Gly Phe Asn Thr Tyr Lys Lys				
	625		630		635
	Gly Leu Phe Thr Gln Asp Asp Val Leu Gly Val Lys Leu Val Gly Tyr				
	645		650		655
30	Arg Ser Phe Ile Lys Asn Tyr Ile His Asn Val Tyr Gly Val Trp Trp				
	660		665		670
	Arg Asp Gly Met Pro Thr Trp Ala Glu Ser Asn Gly Phe Lys Tyr Thr				
	675		680		685
	Ile Ala His Gln Asn Tyr Lys Pro Ile Val Lys Lys Ser Gly Val Glu				
35	690		695		700
	Leu Glu Ile Asn Tyr Asp Met Gly Arg Phe Phe Ala Asn Val Ser Tyr				
	705		710		715
	Ala Tyr Gln Arg Thr Asn Gln Pro Thr Asn Tyr Ala Asp Ala Ser Pro				
	725		730		735
40	Arg Pro Asn Asn Ala Ser Gln Glu Asp Ile Leu Lys Gln Gly Tyr Gly				
	740		745		750
	Leu Ser Arg Val Ser Met Leu Pro Lys Asp Tyr Gly Arg Leu Glu Leu				
	755		760		765

Gly Thr Arg Trp Phe Asp Gln Lys Leu Thr Leu Gly Leu Ala Ala Arg
 770 775 780
 Tyr Tyr Gly Lys Ser Lys Arg Ala Thr Ile Glu Glu Glu Tyr Ile Asn
 785 790 795 800
 5 Gly Ser Arg Phe Lys Lys Asn Thr Leu Arg Arg Glu Asn Tyr Tyr Ala
 805 810 815
 Val Lys Lys Thr Glu Asp Ile Lys Lys Gln Pro Ile Ile Leu Asp Leu
 820 825 830
 His Val Ser Tyr Glu Pro Ile Lys Asp Leu Ile Ile Lys Ala Glu Val
 10 835 840 845
 Gln Asn Leu Leu Asp Lys Arg Tyr Val Asp Pro Leu Asp Ala Gly Asn
 850 855 860
 Asp Ala Ala Ser Gln Arg Tyr Tyr Ser Ser Leu Asn Asn Ser Ile Glu
 865 870 875 880
 15 Cys Ala Gln Asp Ser Ser Ala Cys Gly Gly Ser Asp Lys Thr Val Leu
 885 890 895
 Tyr Asn Phe Ala Arg Gly Arg Thr Tyr Ile Leu Ser Leu Asn Tyr Lys
 900 905 910
 Phe

20

<210> 3

<211> 2757

<212> DNA

25

<213> non typeable *Haemophilus influenzae* strain 3224

<400> 3

atgaagaaag ctataaaatt aaatttaatt acacttagcc taatcaatac aatcggtatg 60
 acgattacac aagctcaagc cgaagaaaca ttagggcaaa ttgatgtcgt agaaaaagtg 120
 30 atatcaaatg acaaaaaacc ttacttgaa gccaaagcca aaagtacgcg tgaaaatgct 180
 ttaaggaaa cacaaaccat tgaccaagtc attcggagca ttctggggc atttactcaa 240
 caagataaag gctcgggtgt ggtttctgta aatattcgtg gcgaaaatgg attaggtcgt 300
 gtcaatacga tggttgatgg tgtaacccaa accttctatt ctacagcctt agactctggt 360
 caatcaggcg gaagttctca atttggtgctg gcaatcgacc ctaattttat tgcagggtgta 420
 35 gatgttaata aaagcaactt ttcgggagca agcgggtataa atgccttagc aggcagtgtc 480
 aatttttagaa cattaagcgt taatgatgtg attaccgatg acaaaccatt cggcattatt 540
 ctgaaaggaa tgacaggag caatgccact aaatccaatt ttatgacgac agctgcaggc 600
 agaaaatggc ttgataatgg tggctatgta ggcgtagtgt atggttatag ccaacgtgaa 660
 gtttcacaag attatcgtat aggtggcgga gaacgattag catcattagg gcaagatatt 720
 40 cttgctaaag aaaaagaaaa gatttttcgt aatgatggtt atgttttaaa ttctgtctgga 780
 caatgggcac ctgatttaaa caaacacat tggctctgta ataccccgag ttctttaaaa 840
 gataaaagta tgagtacatc ttgtaagcct tatcgtcttg gacctgctgc aacgactaga 900
 caagaaattc taaaagaatt attagaagat ggaaaagaac ctaaggatat tgaaaagctc 960

```

caaaaaagta atgatggaat tgaagaaact gaaaaatcat ttgaacgtaa taaagatcaa 1020
tatgacgtcg cccctattga gcctggtagt ttgcaatctc gtacacgtag tcatttatta 1080
aaatttgaat atagcgatga tcaccatacg ctaggggcgc aaatacgtac ccttgataat 1140
aaaattgggt ctcgcaaaat tgaaaaccgt aattaccaag tcaattataa cttcaataat 1200
5 aacagctatc ttgatcttaa tttaatggct gcacataaca ttggcaaaac tatttatcct 1260
aagggtgggt tttttgctgg ctggcaagtg gcagacaaac ttatcacaaa aaatgtggca 1320
aatattgttg atataaataa cagccatact ttcttactgc caaaagaaat cgatttaaaa 1380
accacattag ggtttaacta ttttaccat gaatacagta aaaaccgttt tccagaagaa 1440
ttaagtttgt tttatgtgaa tgaatcacat gatcaaggct tatattcact cagtaataaa 1500
10 gggcgatatt ctggctcaaa aggtttatta ccacaacgtt cagtaatctt acaaccttct 1560
ggcaagcaaa aattttaaac agtgtatttt gataccgcac tttctaaagg tatttatcat 1620
ttaaattaca gcgtgaattt taccattat gcctttaatg gtgagtatgt tggatataaa 1680
aatacagcag ataaaaattaa tgaacctatt ttgcataaat cagggcataa aaaggcatte 1740
aatcattctg ctacattaag tgcagagcta agtgattatt ttatgccatt ttttacttat 1800
15 tcacgcacac acagaatgcc gaatattcaa gagatgtttt tctctcaagt gtctgatgct 1860
ggggtaaaac ccgcattaaa acctgaacaa tctgacacct atcaactagg ctttaatact 1920
tataaaaaag gtctattcac tcaagacgat gtattaggca tcaaattagt gggctatcgt 1980
agctttatta aaaactatat ccacaatgtg tatggagatt ggtcacgaga tgggtttatg 2040
ccagagtggg caagactcaa tgggttttcgt ctgacgattg ctcacaaaaa ttatcaacca 2100
20 atagtgaata aaagcggagc tgagtttagag ctcaattatg atatggggcg tttttttgca 2160
aatctgtctt atgcttatca acgtactaat cagccaacca attatgccga tgccagctca 2220
cgtccgcgta atgcttcaaa agaagagatt ttgaaacaag gttatgggtt atcacgaatc 2280
tctatgttac caaaggacta cggtagatta gagcttggca cacgctgggt tgatcaaaaa 2340
ttaactcttg gtatcgagc ccgttactat ggaaaaagta aacgtgctac aactcaagaa 2400
25 gaatacatca acggctctcg ctatgaaaaa aatactacgc gcgacagaat ttattatgct 2460
attaaaaaga cagaagagat taaaaaacia cctattattt tagatttaca cgtcagctat 2520
gaaccaatca aagatttgat tattaaagcg gaagtacaaa atctattaga taaacgttat 2580
gttgatccgt tagatgctgg aaatgatgag gcttcgcaac gttattattc aagtttaaat 2640
gattcttttag cctgtaaaat aaatgaatca acctgtaatg atggttcaga gaaaactgtg 2700
30 ctttataact ttgcacgtgg aagaacttat attctgagtt tgaactataa attctag 2757

```

<210> 4

<211> 918

<212> PRT

35 <213> non typeable *Haemophilus influenzae* strain 3224

<400> 4

```

Met Lys Lys Ala Ile Lys Leu Asn Leu Ile Thr Leu Ser Leu Ile Asn
1           5           10           15
40 Thr Ile Gly Met Thr Ile Thr Gln Ala Gln Ala Glu Glu Thr Leu Gly
           20           25           30
Gln Ile Asp Val Val Glu Lys Val Ile Ser Asn Asp Lys Lys Pro Phe
           35           40           45

```

	Thr	Glu	Ala	Lys	Ala	Lys	Ser	Thr	Arg	Glu	Asn	Val	Phe	Lys	Glu	Thr	
	50						55					60					
	Gln	Thr	Ile	Asp	Gln	Val	Ile	Arg	Ser	Ile	Pro	Gly	Ala	Phe	Thr	Gln	
	65					70					75					80	
5	Gln	Asp	Lys	Gly	Ser	Gly	Val	Val	Ser	Val	Asn	Ile	Arg	Gly	Glu	Asn	
					85					90					95		
	Gly	Leu	Gly	Arg	Val	Asn	Thr	Met	Val	Asp	Gly	Val	Thr	Gln	Thr	Phe	
					100				105					110			
10	Tyr	Ser	Thr	Ala	Leu	Asp	Ser	Gly	Gln	Ser	Gly	Gly	Ser	Ser	Gln	Phe	
			115					120					125				
	Gly	Ala	Ala	Ile	Asp	Pro	Asn	Phe	Ile	Ala	Gly	Val	Asp	Val	Asn	Lys	
			130					135					140				
	Ser	Asn	Phe	Ser	Gly	Ala	Ser	Gly	Ile	Asn	Ala	Leu	Ala	Gly	Ser	Ala	
	145					150					155					160	
15	Asn	Phe	Arg	Thr	Leu	Ser	Val	Asn	Asp	Val	Ile	Thr	Asp	Asp	Lys	Pro	
					165					170					175		
	Phe	Gly	Ile	Ile	Leu	Lys	Gly	Met	Thr	Gly	Ser	Asn	Ala	Thr	Lys	Ser	
					180				185					190			
20	Asn	Phe	Met	Thr	Thr	Ala	Ala	Gly	Arg	Lys	Trp	Leu	Asp	Asn	Gly	Gly	
			195					200					205				
	Tyr	Val	Gly	Val	Val	Tyr	Gly	Tyr	Ser	Gln	Arg	Glu	Val	Ser	Gln	Asp	
			210				215						220				
	Tyr	Arg	Ile	Gly	Gly	Gly	Glu	Arg	Leu	Ala	Ser	Leu	Gly	Gln	Asp	Ile	
	225					230					235					240	
25	Leu	Ala	Lys	Glu	Lys	Glu	Lys	Ile	Phe	Arg	Asn	Asp	Gly	Tyr	Val	Leu	
					245					250					255		
	Asn	Ser	Ala	Gly	Gln	Trp	Ala	Pro	Asp	Leu	Asn	Lys	Pro	His	Trp	Ser	
			260					265						270			
30	Cys	Asn	Thr	Pro	Ser	Ser	Leu	Lys	Asp	Lys	Ser	Met	Ser	Thr	Ser	Cys	
			275					280					285				
	Lys	Pro	Tyr	Arg	Leu	Gly	Pro	Ala	Ala	Thr	Thr	Arg	Gln	Glu	Ile	Leu	
			290				295					300					
	Lys	Glu	Leu	Leu	Glu	Asp	Gly	Lys	Glu	Pro	Lys	Asp	Ile	Glu	Lys	Leu	
	305					310					315					320	
35	Gln	Lys	Ser	Asn	Asp	Gly	Ile	Glu	Glu	Thr	Glu	Lys	Ser	Phe	Glu	Arg	
					325					330					335		
	Asn	Lys	Asp	Gln	Tyr	Asp	Val	Ala	Pro	Ile	Glu	Pro	Gly	Ser	Leu	Gln	
					340				345					350			
40	Ser	Arg	Ser	Arg	Ser	His	Leu	Leu	Lys	Phe	Glu	Tyr	Ser	Asp	Asp	His	
			355					360					365				
	His	Thr	Leu	Gly	Ala	Gln	Ile	Arg	Thr	Leu	Asp	Asn	Lys	Ile	Gly	Ser	
			370				375					380					
	Arg	Lys	Ile	Glu	Asn	Arg	Asn	Tyr	Gln	Val	Asn	Tyr	Asn	Phe	Asn	Asn	

	385		390		395		400
	Asn Ser Tyr Leu Asp	Leu Asn Leu Met	Ala Ala His Asn Ile Gly Lys				
		405	410		415		
5	Thr Ile Tyr Pro Lys Gly Gly Phe Phe	Ala Gly Trp Gln Val Ala Asp					
		420	425		430		
	Lys Leu Ile Thr Lys Asn Val Ala Asn Ile Val Asp Ile Asn Asn Ser						
		435	440		445		
	His Thr Phe Leu Leu Pro Lys Glu Ile Asp Leu Lys Thr Thr Leu Gly						
		450	455		460		
10	Phe Asn Tyr Phe Thr Asn Glu Tyr Ser Lys Asn Arg Phe Pro Glu Glu						
		465	470		475		480
	Leu Ser Leu Phe Tyr Val Asn Glu Ser His Asp Gln Gly Leu Tyr Ser						
		485	490		495		
	Leu Ser Asn Lys Gly Arg Tyr Ser Gly Ser Lys Gly Leu Leu Pro Gln						
15		500	505		510		
	Arg Ser Val Ile Leu Gln Pro Ser Gly Lys Gln Lys Phe Lys Thr Val						
		515	520		525		
	Tyr Phe Asp Thr Ala Leu Ser Lys Gly Ile Tyr His Leu Asn Tyr Ser						
		530	535		540		
20	Val Asn Phe Thr His Tyr Ala Phe Asn Gly Glu Tyr Val Gly Tyr Lys						
		545	550		555		560
	Asn Thr Ala Asp Lys Ile Asn Glu Pro Ile Leu His Lys Ser Gly His						
		565	570		575		
	Lys Lys Ala Phe Asn His Ser Ala Thr Leu Ser Ala Glu Leu Ser Asp						
25		580	585		590		
	Tyr Phe Met Pro Phe Phe Thr Tyr Ser Arg Thr His Arg Met Pro Asn						
		595	600		605		
	Ile Gln Glu Met Phe Phe Ser Gln Val Ser Asp Ala Gly Val Asn Thr						
		610	615		620		
30	Ala Leu Lys Pro Glu Gln Ser Asp Thr Tyr Gln Leu Gly Phe Asn Thr						
		625	630		635		640
	Tyr Lys Lys Gly Leu Phe Thr Gln Asp Asp Val Leu Gly Ile Lys Leu						
		645	650		655		
	Val Gly Tyr Arg Ser Phe Ile Lys Asn Tyr Ile His Asn Val Tyr Gly						
35		660	665		670		
	Asp Trp Ser Arg Asp Gly Val Met Pro Glu Trp Ala Arg Leu Asn Gly						
		675	680		685		
	Phe Arg Leu Thr Ile Ala His Gln Asn Tyr Gln Pro Ile Val Lys Lys						
		690	695		700		
40	Ser Gly Ala Glu Leu Glu Leu Asn Tyr Asp Met Gly Arg Phe Phe Ala						
		705	710		715		720
	Asn Leu Ser Tyr Ala Tyr Gln Arg Thr Asn Gln Pro Thr Asn Tyr Ala						
		725	730		735		

	Asp	Ala	Ser	Ser	Arg	Pro	Arg	Asn	Ala	Ser	Lys	Glu	Glu	Ile	Leu	Lys	
						740					745					750	
	Gln	Gly	Tyr	Gly	Leu	Ser	Arg	Ile	Ser	Met	Leu	Pro	Lys	Asp	Tyr	Gly	
						755					760					765	
5	Arg	Leu	Glu	Leu	Gly	Thr	Arg	Trp	Phe	Asp	Gln	Lys	Leu	Thr	Leu	Gly	
						770										780	
	Ile	Ala	Ala	Arg	Tyr	Tyr	Gly	Lys	Ser	Lys	Arg	Ala	Thr	Thr	Gln	Glu	
						785										800	
	Glu	Tyr	Ile	Asn	Gly	Ser	Arg	Tyr	Glu	Lys	Asn	Thr	Thr	Arg	Asp	Arg	
10						805										815	
	Ile	Tyr	Tyr	Ala	Ile	Lys	Lys	Thr	Glu	Glu	Ile	Lys	Lys	Gln	Pro	Ile	
						820										830	
	Ile	Leu	Asp	Leu	His	Val	Ser	Tyr	Glu	Pro	Ile	Lys	Asp	Leu	Ile	Ile	
						835										845	
15	Lys	Ala	Glu	Val	Gln	Asn	Leu	Leu	Asp	Lys	Arg	Tyr	Val	Asp	Pro	Leu	
						850										860	
	Asp	Ala	Gly	Asn	Asp	Ala	Ala	Ser	Gln	Arg	Tyr	Tyr	Ser	Ser	Leu	Asn	
						865										880	
	Asp	Ser	Leu	Ala	Cys	Lys	Ile	Asn	Glu	Ser	Thr	Cys	Asn	Asp	Gly	Ser	
20						885										895	
	Glu	Lys	Thr	Val	Leu	Tyr	Asn	Phe	Ala	Arg	Gly	Arg	Thr	Tyr	Ile	Leu	
						900										910	
	Ser	Leu	Asn	Tyr	Lys	Phe											
						915											

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 00/01423

A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/31 C07K14/285 C07K16/12 A61K31/70 A61K39/102 A61K39/40 G01N33/50		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N C07K A61K G01N		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	FLEISCHMANN R. ET AL.: "Haemophilus influenzae Rd section 116; AC U32801; P45114" EBI DATABASE, 1995, XP002144210 the whole document	1-13
X	WO 96 33276 A (HUMAN GENOME SCIENCES INC ;UNIV JOHNS HOPKINS (US)) 24 October 1996 (1996-10-24)	1-13
A	see description pp.1-46; p.76.30	14-16
A	WO 97 01638 A (CORTECS INTERNATIONAL LIMITED ;KYD JENNELLE (AU); CRIPPS ALLAN (AU) 16 January 1997 (1997-01-16) the whole document	1-13
-/-		
<div style="display: flex; justify-content: space-between;"> <input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex. </div>		
<div style="display: flex;"> <div style="flex: 1;"> <p>* Special categories of cited documents :</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="flex: 1;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"&" document member of the same patent family</p> </div> </div>		
Date of the actual completion of the international search <div style="text-align: center; font-weight: bold;">4 August 2000</div>		Date of mailing of the international search report <div style="text-align: center; font-weight: bold;">17/08/2000</div>
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer <div style="text-align: center; font-weight: bold;">Kania, T</div>

INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 00/01423

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 680 765 A (AMERICAN CYANAMID CO) 8 November 1995 (1995-11-08) the whole document ---	1-13
A	WO 93 19090 A (BARENKAMP STEPHEN J) 30 September 1993 (1993-09-30) the whole document ---	1-13
A	FLEISCHMANN R D ET AL: "WHOLE-GENOME RANDOM SEQUENCING AND ASSEMBLY OF HAEMOPHILUS INFLUENZAE RD" SCIENCE,US,AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, vol. 269, no. 5223, 28 July 1995 (1995-07-28), pages 496-512, XP000517090 ISSN: 0036-8075 cited in the application the whole document -----	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 00/01423

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9633276 A	24-10-1996	AU 5552396 A CA 2218741 A EP 0821737 A JP 11501520 T	07-11-1996 24-10-1996 04-02-1998 09-02-1999
WO 9701638 A	16-01-1997	AU 708061 B AU 6236396 A CA 2225405 A EP 0835313 A JP 11510373 T NO 976119 A NZ 311120 A	29-07-1999 30-01-1997 16-01-1997 15-04-1998 14-09-1999 26-02-1998 28-10-1999
EP 0680765 A	08-11-1995	US 5770213 A AU 704882 B AU 1781495 A CA 2148563 A FI 952144 A JP 8337598 A NO 951750 A	23-06-1998 06-05-1999 16-11-1995 06-11-1995 06-11-1995 24-12-1996 06-11-1995
WO 9319090 A	30-09-1993	AU 3916893 A BR 9306109 A CA 2131837 A EP 0632814 A FI 944273 A JP 2810235 B JP 7506248 T NO 943431 A US 5549897 A US 5603938 A US 5876733 A	21-10-1993 18-11-1997 17-09-1993 11-01-1995 15-11-1994 15-10-1998 13-07-1995 10-11-1994 27-08-1996 18-02-1997 02-03-1999